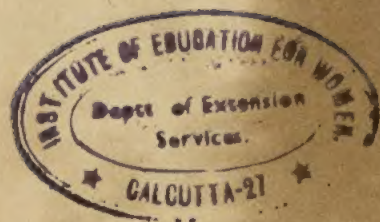


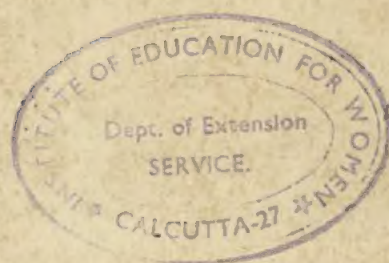
This book was taken from the Library of
Extension Services Department on the date
last stamped. It is returnable within 7 days.

--	--	--	--

VI * CALCUTTA-27 * N31



INTRODUCTION TO TESTING
AND THE USE OF TEST RESULTS
IN PUBLIC SCHOOLS



EDUCATION FOR LIVING SERIES

Under the Editorship of

H. H. REMMERS

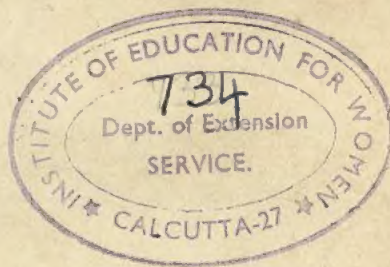
Introduction to Testing and the Use of Test Results in Public Schools

BY

ARTHUR E. TRAXLER
ROBERT JACOBS
MARGARET SELOVER
AGATHA TOWNSEND

with the advice and coöperation of

THE PUBLIC SCHOOLS ADVISORY COMMITTEE
OF THE EDUCATIONAL RECORDS BUREAU



HARPER & BROTHERS, PUBLISHERS
NEW YORK

INTRODUCTION TO TESTING AND THE USE OF TEST RESULTS IN PUBLIC SCHOOLS

Copyright, 1953, by Harper & Brothers. Printed in the United States of America. All rights in this book are reserved. No part of the book may be used or reproduced in any manner whatsoever without written permission except in the case of the cumulative record forms shown and in the case of brief quotations embodied in critical articles and reviews. For information address Harper & Brothers, 49 East 33rd Street, New York 16, N. Y.

EDUCATIONAL RECORDS BUREAU

Incorporated and Chartered by the Board of Regents
of the University of the State of New York

BOARD OF TRUSTEES

FOR COLLEGES AND UNIVERSITIES

Harry J. Carman
Columbia University

Henry T. Moore
Skidmore College

Henry M. Wriston
Brown University

FOR INDEPENDENT SCHOOLS

Hart Fessenden
The Fessenden School

William L. W. Field
National Council of
Independent Schools

Allan V. Heely, Chairman
The Lawrenceville School

FOR PUBLIC SCHOOLS

Howard V. Funk
Bronxville Public School
Bronxville, New York

Galen Jones
U.S. Office of Education

Lester W. Nelson
Scarsdale Public Schools
Scarsdale, New York

PUBLIC SCHOOLS ADVISORY COMMITTEE

Paul Axtell
Caldwell Public Schools
Caldwell, New Jersey

Arthur S. Bibbins
Darien High School
Darien, Connecticut

R. Elizabeth Blaul
Highland Park High School
Highland Park, Illinois

Miriam B. Booth
School District of Erie
Erie, Pennsylvania

Elizabeth Jean Brown
Rye High School
Rye, New York

Irene A. Fike
Scarsdale Public Schools
Scarsdale, New York

Adria Galbraith
Plainfield High School
Plainfield, New Jersey

Frank C. Ketler
Cheltenham Township Schools
Elkins Park, Pennsylvania

Julia A. Markham
Bronxville Public School
Bronxville, New York

Blanche Matthews
Hillside Public Schools
Hillside, New Jersey

Frank Misner
Bronxville High School
Bronxville, New York

Clarence G. Noyce
White Plains Public Schools
White Plains, New York

Frederick M. Raubinger
Ridgewood Public Schools
Ridgewood, New Jersey

Russell H. Rupp
Shaker Heights High School
Cleveland, Ohio

Joseph G. Shuttlesworth
Summit High School
Summit, New Jersey

Edward Summerton
Weston Public Schools
Weston, Connecticut

CHAIRMEN OF COMMITTEES

Public Schools Advisory Committee, *Arthur S. Bibbins*, Darien High School
Independent Schools Advisory Committee, *Hans Froelicher, Jr.*, The Park School, Baltimore
Committee on Tests and Measurements, *J. Wayne Wrightstone*, New York City Public Schools
Committee on School and College Relations, *Burton P. Fowler*, Germantown Friends School

STAFF

Ben D. Wood, Director
Arthur E. Traxler, Executive Director
Agatha Townsend, Research Associate
Geraldine Spaulding, Test Program Supervisor
Victor Nappi, Business Manager



Contents

Foreword	ix
1. A Point of Departure	1
2. What Do Tests Contribute to Understanding the Individual Pupil?	5
3. How Shall We Plan a Testing Program?	13
4. How Can Tests Be Selected?	20
5. How Should Tests Be Given?	30
6. How Should Tests Be Scored?	36
7. How Shall We Analyze and Interpret Test Results?	44
8. How Shall We Record Test Results?	68
9. How Shall We Use Test Results?	89
10. How Does All This Apply to a Specific Case?	96
Appendix	109
Index	111

Foreword

IN 1939, Dr. Margaret Selover, who was a member of the staff of the Educational Records Bureau, prepared some material to provide a simple, nontechnical treatment of testing and the use of test results, with particular reference to independent or private schools. That material was submitted a section at a time to Bureau member schools for criticism. The suggestions received from the schools were taken into account in the preparation of a revision which was put together in loose-leaf binders for distribution.

A thoroughgoing section revision of this material was begun in 1948 by members of the Bureau staff, with the advice and coöperation of members of the Committee on Tests and Measurements. In some sections, the changes were minor, whereas other sections were completely rewritten. That revision, which was largely the work of Dr. Agatha Townsend, stressed the use of test results in independent schools. It was issued by the Educational Records Bureau in an offset edition in 1950 as Educational Records Bulletin No. 55, *Introduction to Testing and the Use of Test Results*.

The 1950 edition was used by many independent schools and by a considerable number of public schools. Various public schools informed the Bureau that they found the material helpful but that they felt that it would be still more helpful if another edition designed especially for teachers and counselors in public schools could be made available. Accordingly, the Bureau's Public Schools Advisory Committee decided to sponsor the preparation of a public school edition of this material. This revision was undertaken by Dr. Robert Jacobs, then a member of the Bureau staff and more recently director of counseling at the Agricultural and Mechanical College of Texas. It was largely his work which made this book on testing in public schools possible. Although a considerable amount of the original material was retained, several chapters were completely rewritten. The revision was read critically by members of the Public Schools Advisory Committee, all of whom were administrators or faculty members of public schools, and further revision was made in the manuscript in accordance with their suggestions. Hence, the final product is an outgrowth of the work of many persons, and it is strongly influenced by the viewpoint and expressed needs of teachers and counselors in a large number of schools. The authors wish to express special appreciation for the coöperation of Professor Herschel T. Manuel of the University of Texas, who read the entire manuscript and made constructive suggestions.

1891

It is hoped that the guidelines will serve as a practical, step-by-step guide for the research community. The list of reference items, the further discussion groups, the exercises, training programs, for parents, who have had experience with tests but who desire to become up to the simpler instruments of testing, and the introductory classes in this and the above-mentioned.

New York City
(October, 1952)

Литвиш Е. Троцкий

INTRODUCTION TO TESTING
AND THE USE OF TEST RESULTS
IN PUBLIC SCHOOLS



A Point of Departure

(7) *admission to nursing school*
 by a person who is a citizen
 of the United States, or a person
 who is a permanent resident of the
 United States, or a person who is

[illegible]

[Faint, illegible text]

[illegible]

INTRODUCTION TO TESTING

The needs issuing from emphasis upon individualized education are expressed more concretely in a report of the American Association of School Administrators' Commission on Youth Problems. This report, appearing in the Sixteenth Yearbook of the association, stated four fundamental needs:

1. Each child should be entitled to from ten to fifteen years of day school instruction without the humiliation of repeated failures and retardation; should follow a school program, adapted to his abilities and interests, thru which he may achieve a reasonable measure of school success; and should have the right to associate in school with others of his own age and degree of physical and social maturity in activities thru which he may develop desirable social skills and a wholesome personality.

2. Each child should have assistance in overcoming any individual handicaps, or in learning to face them frankly and courageously; in discovering and developing any special abilities that he may have in becoming acquainted with educational and occupational opportunities which are in harmony with his abilities, interests, ambitions, and prospects; and in making wisely the choices leading toward an occupation. Such orientation in occupational, economic, and social problems is a *basic part of general education* and is fundamental to all special guidance services.

3. Each child should have an opportunity in school not only to choose his occupation but to begin his preparation for occupational life and to develop initial marketable skills. He should have assistance, if necessary, in securing employment in a suitable occupation and in making plans for further education to insure growth and advancement in service. If, thru accident or circumstances beyond his control, the skills and abilities which he has developed are no longer marketable, then the school system should provide the necessary guidance and assistance in retraining in order that transfer to some related field may be accomplished in which a reasonable measure of success may be

possible. This involves clinical services to prevent personal unhappiness as well as occupational maladjustment.

4. If a person is unable to achieve self-support and independence because of mental, physical, or personality handicaps, which in the present state of learning may not be overcome, society must provide, in the years immediately before and after school-leaving, special services of guidance and supervision. This may be in the form of special placement and supervision (a) in private employment, (b) in a sheltered workshop of some social agency, or (c) in an institution. Such persons must be protected from exploitation, antisocial or criminal influences, and the dangers of disease and poverty. Each community should provide for continuous study of its potential social problems.¹

Although lip service is given to such concepts of individual differences as those expressed by John Locke and while our teacher-training institutions preach the need for individualized education such as that described in the report of the Commission on Youth Problems, educational practice violates these principles at many points. As Bernard I. Bell wrote in a widely read education issue of *Life* magazine, "Our school system . . . seems to pre-suppose, that for education to be democratic, every man's child must be treated as an equal of every other man's child, both in kind of brains and educability."² In so far as democratic education has been interpreted in this way, planning has been in terms of mass education, and practices have been established which are at variance with a philosophy of individualized education. This gap between philosophy and practice comes not so much from lack of zeal or intent as from practical limitations and difficulties. The public school particularly

¹ *Youth Education Today*, Washington, American Association of School Administrators, Sixteenth Yearbook, 1938, pp. 173-174.

² Bernard I. Bell, "Know How vs. Know Why," *Life*, 29: 89-92, 97-98 (October 16, 1950).

A POINT OF DEPARTURE

finds itself beset with problems which are yet to be solved.

Many factors have contributed to the current difficulties of the public schools. During the past half-century or so, the population of the United States has doubled, but the number of students enrolled in the nation's high schools has been multiplied by ten. The overwhelming increase in enrollment was brought about largely by a changing economic pattern. Truancy laws and child labor laws sent a greater percentage of the nation's children into the schools. During the depression years which preceded the last war there was little employment to be found by the young person. The alternative of attending high school was a logical choice. Too, the importance of high-school education in vocational advancement and occupational success has increased with the gradual breakdown of areas of work into specialized pursuits.

The increasing complexity of the world of work presents another problem to the public school. The amount and range of information needed by youth has increased considerably. No longer are the "three R's" sufficient for effective living. The importance of tool subjects has never diminished, but the schools have been called upon to add courses oriented toward vocational training. New demands have been made in the areas of science and social studies. Other social institutions are looking to the school for assumption of responsibility with regard to character development, sex education, wise use of leisure time, and many other phases of living.

The listing of influences which have brought about the present dilemma is by no means exhausted. However, it is not the purpose of this book to delve into the history of education. The

preceding brief discussion is intended simply to argue a point for the harassed public-school teacher and to indicate that the very factors which emphasize the need for better understanding of the individual pupil operate to make the process of understanding more difficult. Since the public-school teacher, in dealing with large groups of pupils, cannot hope to realize anything approaching the Hopkins-Garfield relationship with each pupil, techniques, tools, and methods are needed which will bring as much individualization as possible into the classroom situation.

Individual differences among boys and girls cannot easily be identified by observation alone. To find out what the pupil brings to the learning situation, how far he may be expected to go, what direction he may take, and what difficulties he may encounter, observation must be supplemented by more searching and more objective techniques. Within the past half-century new methods of analysis and appraisal have been developed to assist in understanding the child. Among these are anecdotal records, projective techniques, rating scales, and objective measurement.

The chapters which follow deal with questions commonly asked by teachers who participate for the first time in a program of objective testing. An attempt is made to present basic and essential facts in considering each question. The treatment is meant to serve as an introduction to the subject of reasons for testing, the way to test, and the proper interpretation of test results, with specific orientation to the needs of public schools. Pertinent references for anyone who wishes to read more extensively are appended to each chapter.

INTRODUCTION TO TESTING

SUGGESTIONS FOR FURTHER READING

1. Donahue, Wilma T., Coombs, Clyde H., and Travers, Robert M. W. (eds.), *The Measurement of Student Adjustment and Achievement*, Ann Arbor, Michigan, University of Michigan Press, 1949, pp. 3-22.
2. Erickson, Clifford E., *A Basic Text for Guidance Workers*, New York, Prentice-Hall, Inc., 1947, pp. 17-35.
3. Froehlich, Clifford P., *Guidance Services in Smaller Schools*, New York, McGraw-Hill Book Company, 1950.
4. Hamrin, Shirley A., and Paulson, Blanche B., *Counseling Adolescents*, Chicago, Science Research Associates, 1950, pp. 111-142.
5. Remmers, H. H., and Gage, N. L., *Educational Measurement and Evaluation*, New York, Harper & Brothers, 1943, pp. 19-37.
6. Ross, C. C., *Measurement in Today's Schools*, New York, Prentice-Hall, Inc., 2nd ed., 1947, pp. 3-26.
7. Rothney, J. W. M., and Roens, B. A., *Guidance of American Youth*, Cambridge, Massachusetts, Harvard University Press, 1950.
8. Traxler, Arthur E., *Techniques of Guidance*, New York, Harper & Brothers, 1945, pp. 1-13.
9. Wood, Ben D., and Haefner, Ralph, *Measuring and Guiding Individual Growth*, New York, Silver Burdett Company, 1948, pp. 513-523.

What Do Tests Contribute to Understanding the Individual Pupil?

A NEW teacher enters the classroom to meet her first class. She sees an array of faces—the boys and girls who are to share learning experiences under her guidance throughout the school year. Each represents an unknown area which she must explore if she is to do an effective job in this guidance process. She may begin by learning the names; this will be a relatively simple task. A few other aspects of individuality may be learned perhaps as easily, but the teacher knows that beneath the outward appearance of these faces the basic elements which make one pupil different from another are too well concealed to yield to informal observation alone. How, then, can the teacher come to know and understand the individual pupil? Or perhaps we should ask first: What kinds of information will assist the teacher in understanding the pupil?

For adequate guidance and instruction information is needed in a number of areas. They may for convenience be grouped into two general questions: (1) Where does the pupil now stand with respect to abilities, interests, achievement, and personal and social adjustment? and (2) How far and in what directions can he be

expected to go in terms of his capacities, limitations, and needs?

To answer the first question, information is needed concerning the level of the pupil's general ability and the nature of any special abilities he possesses. The knowledge and skills he has acquired through both school and out-of-school experiences form a part of this "present status" picture. Also, much of what he is *now* is determined by his relations to others, his initiative, his feelings of security, the degree of self-confidence he displays, and other elements which enter into personal and social adjustment. Information in all of these areas is needed to determine just what the child brings to the learning situation at the outset.

The second question concerns rate, ceiling, and direction of growth. Here information is needed regarding basic interests of the individual, the kinds of goals he has set for himself, and the appropriateness of these goals in terms of his general ability and special aptitudes. Actually, prediction and judgment are involved in answering this question. It is essential, though, that this projection be made if the teacher is to assist the individual in maximum

INTRODUCTION TO TESTING

fulfillment of his particular capacities and the direction of energies toward attainable life goals.

From this general treatment can be drawn some five areas of information which have particular significance in understanding the individual. They are:

1. General aptitude or ability.
2. Special aptitudes and abilities.
3. Achievement in different fields of study.
4. Educational and vocational interests.
5. Personal and social adjustment.

Let us see now how tests contribute to each of these five areas.

GENERAL APTITUDE

Great progress has been made in the measurement of aptitude or general mental ability. This type of testing received its first great impetus when Binet found that there were certain tasks (or test situations) which could discriminate between children in the schools of France who would subsequently advance in school at a normal rate and those who would have to be retarded. For some time thereafter in the United States most mental testing was done in connection with the discovery of such retarded children and the admission of the most markedly subnormal into institutions for the feeble-minded.

During the First World War the need of the armed forces for a scientific basis of classifying men inducted into the service brought about great progress in the development of mental tests. After the war mental testing proceeded on almost a wholesale basis. School people became "mental test" conscious; colleges began to test their entering classes; and state-wide school testing programs were organized. Business and industrial organizations took up the idea, enter-

ing enthusiastically into the new fad without sufficient attention to appropriateness or usefulness of the technique for particular situations. The testing movement actually received a temporary setback as a result of this blind enthusiasm. However, during this time and on into the late 1920's and the 1930's, intelligent use of mental tests in the more conservative educational guidance programs tended to offset some of this loss. Here mental measurement found a natural and indispensable place in the educative process. Paralleling the growth of educational guidance and perhaps motivating it in many instances was the application of mental tests in the large-scale counseling services set up by the United States Employment Service during the depression years. World War II gave great impetus to mental tests as well as to all kinds of testing, since there was once more need to classify and use the abilities of large numbers of persons as quickly and effectively as possible. With rapid development in the postwar period, mental testing forms an important part of the systematic programs of objective testing adopted in schools of all kinds throughout the United States.

The general ability or general aptitude test provides an estimate of the intelligence of the individual. Ideally, the mental ability or "intelligence" test presents materials and situations which are new to the testee. His success in dealing with these "new" materials is considered to be indicative of his ability to think and act intelligently in new situations. However, it is virtually impossible to develop test items which are *new* in all respects. Hence, past experience and learning are usually reflected in the results of tests of mental ability. Frequently, these tests are designed to predict the individual's capacity to cope with the school curriculum. Hence,

WHAT DO TESTS CONTRIBUTE TO UNDERSTANDING THE INDIVIDUAL PUPIL?

they are sometimes called scholastic aptitude tests.

Mental ability tests may be grouped in a number of ways. They may be classified, for example, with regard to the method of administration, some being group tests, some designed for administration to one examinee at a time. They may be grouped according to type of items, either verbal or nonverbal, the former depending upon language symbols as a medium of test performance, the latter dealing with pictures, figures, and other nonverbal symbols. A third grouping could be made on the basis of results yielded, some tests providing a description of intelligence in terms of a single score or "I.Q.," others yielding a description in terms of a profile of separate mental traits, such as verbal ability, number ability, spatial relations, reasoning, and so forth.

Obviously, the mental ability test supplies valuable information toward understanding the individual pupil. This information concerns the general mental level of the individual and the differential aspects of mental make-up, such as number, spatial, verbal, and so forth.

SPECIAL APTITUDES

Development of objective techniques for the measurement of special aptitudes was a natural outgrowth of the testing movement. Its history is quite similar to that of mental testing. Experiment and research with various techniques during the First World War and during guidance and counseling applications of testing between the two wars, followed by more widespread application of objective measurement techniques in World War II, form the general background from which this type of test emerged. Special aptitude tests are sometimes confused with dif-

ferential ability tests, such as those mentioned in the preceding section as designed to describe intelligence in terms of separate components. Confusion probably is due to inadequacy of current test terminology in expressing the *functional* difference between the two types of instruments. The differential ability test, which describes intelligence in the form of a profile of mental abilities, aims for the identification of separate factors, whereas the special aptitude test attempts to measure a combination of factors which may relate to success in special occupational fields, such as medicine, accounting, law, nursing, stenography, or mechanics. In addition, there are tests which predict broadly the individual's possibilities of success in art, music, clerical work, teaching, dentistry, or other vocational pursuits.

The distinction between the aptitude test and the achievement test which is described in the next section is not always clear cut. Achievement tests may be described as those tests which are intended to measure what a pupil has learned. Obviously, the amount a pupil already has learned in a particular field may often be a good basis for predicting how much he will learn in the future in that or related areas. Therefore, achievement tests in many cases may be put to the same uses as aptitude tests. For example, an achievement test in general mathematics given at the end of the eighth grade may serve as a good basis for predicting subsequent success in elementary algebra or even in predicting later vocational success in occupations where mathematics would be applied.

Special aptitude tests, then, contribute information helpful in understanding the individual's capacities and limitations as they may relate to possibilities of success in various fields of endeavor.

INTRODUCTION TO TESTING

ACHIEVEMENT

The academic growth or development of the pupil is usually observed in terms of achievement. What the pupil learns not only reflects the methods, techniques, and content of instruction but also reveals in large measure his interests and abilities. Hence, it is important to follow carefully the rate and direction of subject matter mastery. This type of information is supplied by the achievement test.

Achievement tests are, of course, not new. Teachers have been testing their pupils' knowledge by means of oral quizzes and written examinations almost since schools began. Written examinations of the essay type frequently consisted of ten questions, each worth ten points, or twenty questions, each worth five points. Often they began with "Explain," or "Discuss," or "Trace." After a time an occasional teacher began to wonder, "Would another teacher who didn't know my pupils grade the tests the same way?" Various studies have shown that usually the answer to this question is "No!" For example, one geometry instructor had reproduced a geometry test paper handed in by one of his pupils and sent it to many mathematics teachers with the request that they rate it on a scale of 100 points. The paper came back with grades ranging from 10 to 90. Similar results have been obtained on tests in English and in other fields. The situation can be improved by constructing the questions carefully to avoid all possible ambiguities and by making an elaborate key showing all types of responses and the amount of credit to be given for each item of information. Even so, in most cases these tests must be scored by a teacher, or at least by persons who are thoroughly familiar with the subject matter covered.

The achievement of some instructional objectives is measured rather effectively by the essay examination. It is obvious, for example, that ability to write an essay will be revealed in this type of test situation, and that the attainment of this particular objective—say, in an English course—usually will be evaluated more effectively by the essay question than by the short-answer test, particularly if standards of grading are carefully worked out and applied. Other than in its application to this and to related objectives, such as ability to organize and evaluate broad subject matter areas, the essay test is generally less useful than the more efficient and more reliable method of objective testing. As a result, some teacher-made tests today employ brief, specific questions to each of which only one correct answer will fit. Such short-answer tests reduce the possibility of unreliable grading, and, if scientifically constructed and standardized on representative groups of pupils in the local school, they may become effective instruments for use in the process of understanding the growth and the development of each individual pupil. If the requirements of scientific construction and adequate standardization are met, the teacher-made test may be more effective in some situations than the published objective test, simply because it is constructed with specific, local instructional objectives in mind.

Unfortunately, few teachers and few schools are equipped to do an adequate job of either constructing or standardizing the objective test. Success in test construction involves, in addition to subject matter experience, some knowledge of statistics, of uses and limitations of various types of test items, of techniques of item writing, and of other technical information with which the average teacher usually is unfamiliar. Test standardization, again, requires some special

WHAT DO TESTS CONTRIBUTE TO UNDERSTANDING THE INDIVIDUAL PUPIL?

technical knowledge. Also, even if test standardization is handled adequately on a local basis, pupil comparison is limited; that is, there is no way to compare students in the local school with boys and girls at corresponding academic levels in other schools. To illustrate: Jane comes from an elementary school in a neighboring community, and John comes from the local public grade school. When they enter the same secondary school, the algebra teacher finds that John's record includes a test average of 90 in arithmetic, while Jane has an average of only 77; yet Jane knows far more about arithmetic than John. They were given different sorts of examinations or graded on different standards. Moreover, the teacher's grades assigned to these pupils were affected by the fact that John's class as a whole happened to be rather poor, whereas Jane's class was very good in arithmetic. If the algebra teacher tried to divide the class into fast and slow sections on the basis of such marks in arithmetic, he would misplace both John and Jane, and probably many more in the class as well.

Evidently, then, in most school situations the published objective tests are useful in evaluation of pupil growth and development toward instructional objectives, and individual achievement is another kind of information supplied by tests.

INTERESTS

Another important type of objective instrument supplies a fourth kind of information. This is the test of interests, perhaps more appropriately called the inventory of interests, since the answers are not scored as right or wrong. It is known that information concerning the degree of agreement between individual interests and

those of persons successfully engaged in a given field is definitely valuable in predicting the individual's fitness for that field. The customary form of interest inventory is the standardized questionnaire to which the individual responds in some manner to indicate his preferences. Various scoring procedures are used. Under the more sophisticated procedures, the responses are scored with a variety of scales in each of which the answers are weighted with plus or minus values on the basis of research. Most such interest inventories are based on research in occupations, although some are expressed in more general terms with categories which seem to be important interest areas for persons in many fields of work. The configuration of scores on the interest profile, then, assumes importance in occupational guidance. A single interest test may yield individual scores in from nine to forty or more different areas, all based on the same quantitative scale in order to provide direct comparability from one field to another.

The inventory of interests provides another kind of information which is useful in understanding the individual child.

PERSONAL AND SOCIAL ADJUSTMENT

In a fifth classification of objective instruments we place a group of tests which may be loosely described as *personality tests*. They are designed to measure a variety of traits, habits, and attitudes. The list includes personal and social adjustment, feelings toward school, home, or social groups and toward emotionally tinged or crucial areas, such as war, censorship, capital punishment, and Sunday observance. This list of topics could be extended almost infinitely, but it will tend to indicate the inclusive character of

INTRODUCTION TO TESTING

"personality measurement." One of the main difficulties encountered in the construction and use of such tests is that there is little agreement concerning the definition of personality or the nature of personal qualities. There is considerable evidence that behavior traits used in describing personality are largely relative to the situation. For instance, one child may cheat in an examination and yet never touch the property of another child.

Even if this difficulty is overcome, a serious limitation to the paper-and-pencil test of personality is that it usually attempts to measure a person's behavior by asking him how he behaves in a given situation. Of course, this technique is simpler than observing actual behavior in a wide variety of such situations, but at the same time it allows the individual to answer the questions as he thinks they should be answered, or as he wishes he behaved, rather than as he really would behave. Some of this difficulty can be overcome by explaining carefully that the test is not scored for *right* or *wrong* answers and that the test results will be useful only if the pupil is entirely truthful. In other instances, the purposes of the test may be somewhat disguised or the test questions may have an internal check on consistency of responses. This kind of check is sometimes obtained by presenting essentially the same question in slightly different ways in different parts of the test. At any rate, the teacher using such tests must always realize the possibility of the pupil's being influenced by his unconscious desire to put himself in the best possible light.

Because of some of these shortcomings of the paper-and-pencil tests, recent years have seen the development of different means of personality appraisal. Some of these are known as projective techniques. Since they are set up to en-

courage the person examined to put his own interpretation on the test situation, theoretically the individual projects or reveals something about his own personality structure by the reaction he makes to the test items, usually in the unstructured form of ink blots, clouds, pictures, and so on. At the present time, the interpretation of responses on projective tests depends to a marked degree on the training and the insight of the examiner. Hence, these tests are in most instances not "objective" in the sense that this term is used in our discussion.

Personality tests may sometimes be used to stimulate pupils to evaluate their own characteristics, to locate pupils who are poorly adjusted and in need of help, and to serve as a point of departure in interviews with pupils. Suppose, for example, that a pupil with a high aptitude rating and satisfactory previous achievement should be found to be failing, defying his teachers, and bullying younger pupils. The causes for his behavior might be found only after several interviews with many questions about all sorts of situations. However, if certain test results show him to be poorly adjusted with regard to his home and possessed of many feelings of insecurity and loneliness, the teacher or counselor might have reason to suspect a recent upsetting influence in the home and to direct the interview toward that area. A test should never supplant the interview, but it may serve as a point of departure for the discussion of pupil and teacher or counselor and may suggest questions which ought to be considered.

Although there are many limitations to the use of so-called personality tests, their application with caution and intelligence will provide a fifth kind of information to supplement general ability, achievement, aptitude, and interest data supplied by other measures.

WHAT DO TESTS CONTRIBUTE TO UNDERSTANDING THE INDIVIDUAL PUPIL?

This brief review indicates that objective tests form a useful and important device in the process of understanding the child and that they yield objective appraisal to supplement teacher judgment in many areas where information is needed. However, we must not argue the case for objective tests to the point where they may be considered the final answer to the problem of obtaining information for individualized education. It should be pointed out, with considerable emphasis, that tests are subject to a number of limitations, which should be clearly understood. Unfortunately, a true sense of humility with regard to objective testing comes ordinarily only after considerable experience with the device. The beginner is frequently filled with such enthusiasm that proper perspective is lost in the appraisal procedure, and broad generalizations or important decisions may be made on the basis of too little information. Limitations will be discussed in more detail in later chapters, but let us summarize briefly at this point some of the important limitations of the technique of objective measurement:

1. There are important aspects of human behavior as well as important instructional objectives which cannot be evaluated effectively by objective tests available at the present time.

2. Test results are influenced significantly by factors such as motivation, physical condition, and emotional tone, which are often inadequately controlled in the test situation.

3. One is frequently misled by operation of unrecognized factors in testing, e.g., the reading-comprehension factor in arithmetic problem-solving tests, the rate-of-perception factor in closely timed tests, or the general-intelligence factor in achievement testing.

4. Tests must be employed within the limits of the accuracy and consistency with which

they measure whatever they are supposed to measure. No test is perfectly reliable, and practically all tests compromise with regard to validity. The meaning of these two terms will be discussed in Chapter 4.

5. In the main, objective tests are used to describe performance in terms of comparisons with other individuals. This fact may discourage consideration of the pupil within the framework of his own individual capacities, limitations, and goals. As yet, we do not seem to have adequate statistical techniques for describing test performance in terms of individual maturation units.

6. Objective testing is criticized frequently as being atomistic—that is, as approaching an understanding of the child by searching for bits or parts of behavior which are put together to produce a "whole" personality. In at least partial support of this criticism, it must be recognized that human behavior in many situations is meaningful and understandable only in terms of the total personality in a total situation.

7. Closely related to the limitation just given is that of overemphasis on objectivity, which the device claims as its chief advantage. Individual judgment cannot be ruled out of the appraisal process. Even after "facts" are obtained by objective means, there remains the task of fitting them together. This involves judgment, intuition, and discrimination, processes which are *subjective* more often than *objective*.

8. A test score represents a sort of spot check, indicating the individual's status with regard to a particular quality or capacity at a given point in his growth cycle. Since individuals vary with respect both to rate and to ceiling of growth, it is necessary to apply frequent comparable checks in order to obtain an adequate understanding of the individual. One should

INTRODUCTION TO TESTING

be very cautious in generalizing on the basis of a single test result.

SUMMARY

From a background of application in many areas, the technique of objective testing emerges as an important and helpful device to assist the teacher in bringing the educational program

into harmony with individual capacities and goals. At least five kinds of information are supplied by objective tests. However, these instruments have definite limitations, and they should be applied with these limitations in mind. Nonetheless, used with proper perspective and caution, test results form one of the most important sources of personal data basic to the process of individualizing education.

SUGGESTIONS FOR FURTHER READING

1. Bell, John E., *Projective Techniques*, New York, Longmans, Green and Company, 1948.
2. Cronbach, Lee J., *Essentials of Psychological Testing*, New York, Harper & Brothers, 1950, pp. 13-19.
3. Donahue, Wilma T., Coombs, Clyde H., and Travers, Robert M. W. (eds.), *The Measurement of Student Adjustment and Achievement*, Ann Arbor, Michigan, University of Michigan Press, 1949, pp. 51-58.
4. Froehlich, Clifford P., and Benson, Arthur L., *Guidance Testing*, Chicago, Science Research Associates, 1948, pp. 1-4.
5. King, Joseph E., Jr., "Using Tests in the Modern Secondary School," *Bulletin of the National Association of Secondary School Principals*, 32:5-92 (December, 1948).
6. Lindquist, E. F. (ed.), *Educational Measurement*, Washington, American Council on Education, 1950, pp. 3-46, 495-530.
7. Micheels, W. J., and Karnes, M. R., *Measuring Educational Achievement*, New York, McGraw-Hill Book Company, 1950, pp. 22-78.
8. Remmers, H. H., and Gage, N. L., *Educational Measurement and Evaluation*, New York, Harper & Brothers, 1943, pp. 1-12.
9. Ross, C. C., *Measurement in Today's Schools*, New York, Prentice-Hall, Inc., 2nd ed., 1947, pp. 27-64.
10. Sims, Verner M., "The Essay Examination as a Projective Technique," *Educational and Psychological Measurement*, 8:15-31 (Spring, 1948).
11. Sims, Verner M., "Questioning Some Assumptions Underlying Current Achievement Testing," *Educational and Psychological Measurement*, 8:567-573 (Winter, 1948).
12. Wood, Ben D., "The Need for Comparable Measurement in Individualizing Education," *The Educational Record*, 20:14-31 (January, 1939), Supplement No. 12.

How Shall We Plan a Testing Program?

ANY TEST service organization receives hundreds of letters each year containing a variety of questions concerning testing and the use of test results. Not infrequently, the problems take some such form as "I have decided to give some tests in my school this fall. What ones would you recommend?" or perhaps, "We have given a standardized test to all pupils in our school. What are some of the ways in which we can use the results?" Such questions may be taken as evidence that the technique of objective testing is being considered or has been employed without definite planning. To be effective, a testing program must be set up with system and order and with the coöperation and support of all parties concerned. These essentials can be realized only through careful planning.

Test program planning is sometimes undertaken coöperatively, when some agency or commission provides the necessary leadership to encourage groups of schools having fairly common testing needs to work together. The work of one testing and service organization, the Educational Records Bureau, can be cited as an example. Schools participating in the Bureau's

regular spring and fall testing programs are relieved of many of the major responsibilities in planning. The somewhat hazardous task of test selection is performed by the Subcommittee on Test Selection of the Committee on Tests and Measurements. The parent group is made up of testing and guidance personnel appointed from the Bureau membership institutions. Thus, test selection is oriented as nearly as possible to the needs of the member group and is performed in a scientific manner. Many other details of planning, such as establishment of testing dates, preparation of general directions, and scoring and reporting of results, are performed by the Bureau staff.

Since the private, or independent, type of school makes up the major portion of Bureau membership, the regular testing programs reflect notably the needs and objectives of the private-school group. This program is useful and appropriate for many of the public schools holding membership in the Bureau, particularly in those instances where large proportions of college preparatory students are found in the total enrollment. However, for pupils in public schools who are not preparing for college the

INTRODUCTION TO TESTING

regular Bureau program needs to be supplemented or partially replaced by other tests. In some areas, test planning for public schools has been undertaken coöperatively through state or regional commissions. In the main, though, responsibility for providing system and planning to the testing scheme has rested with the individual public school.

It is recognized that no master plan can be set up which will guarantee an effective testing program in all situations. However, certain essentials or characteristics of adequate planning can be stated which will serve as a sort of guide. At the same time, certain steps are common to all testing programs, and these can be listed in the order usually followed.

GENERAL CONSIDERATIONS

In order for a plan of testing to be adequate, it should meet fairly well the following specifications:

1. The tests employed in the program should be selected and administered for specific purposes which are stated in advance.
2. The program should be undertaken coöperatively by the school faculty.
3. A comprehensive list of the procedures involved in carrying out the program must be included in the overall plan.
4. The program should be practical and definite.
5. The program should be continuous and long-range in scope.

The first item in this list of essentials is related closely to the need for stating educational objectives in the local situation. In defining the educational philosophy for a particular school, in setting up the broad general goals toward which changes in pupil behavior are to be di-

rected, and in defining specific instructional objectives for various subject matter areas, there are needs for measurement and evaluation. To meet these needs the testing program should provide information to assist in attainment of goals as well as checks to find if goals have been reached. Some of the needs are met more adequately by methods and techniques other than objective testing. Those items which can be served by test results should be identified and stated clearly in planning the testing program.

It follows that in many instances the initial step in undertaking a testing program will be to define and list the objectives toward which the local educational effort is directed. This procedure will involve a statement from each teacher, or a joint statement from all the teachers in a department, regarding aims and goals for specific courses of study. It will involve, also, group thinking which results in a statement of general goals drawn from community needs and a definition of the overall purpose or philosophy forming the framework within which progress toward such goals will be accomplished. From these statements, then, the particular areas where objective testing will make a contribution are identified.

The second essential listed above deserves particular attention. Actually, the success or failure of the testing program may be determined entirely by the degree to which it grows out of the coöperative effort of the whole school staff. This principle has its counterpart in all activities involving human relations. The industrial psychologist, for example, knows that no new program, such as job evaluation or a wage incentive plan, can be expected to succeed unless it is understood and accepted by the employees. His first step is to set up a working committee and to carry on a program of orien-

HOW SHALL WE PLAN A TESTING PROGRAM?

tation and education to acquaint all departments in an organization with the benefits which will result from a new procedure.

To carry on effectively a long-term testing and guidance program, responsibility must be centered with some person or group of persons technically prepared to administer this particular aspect of the school's work. Nonetheless, if the program is set up and handed down by the specialist, by the school administrator, or by the research department, full coöperation cannot be expected. The entire staff should have an opportunity to contribute to overall planning, indicating individual needs and points of view. In large schools where total group participation would be unwieldy, planning can be carried out by a representative committee and the recommendations then offered to the total group for discussion and approval, so that the testing program is an enterprise in which nearly all will share. It is simply a common-sense principle of human relations that in any endeavor the degree of individual acceptance follows from the degree of understanding and appreciation which comes from sharing in the planning therefor.

To make the school testing program a co-operative endeavor usually it is necessary to prepare the staff with respect to test administration, scoring, and use of results. Again, acceptance and coöperation cannot be expected if the teacher is unfamiliar with procedures and techniques, with meaning of technical terms, or with elementary uses and misuses of test results. Preparation, as a means of understanding and coöperation, can be achieved through study groups and other in-service training efforts. Staff preparation is further emphasized in later chapters dealing with specific aspects of operating the program.

The third characteristic of adequate test pro-

gram planning mentioned above involves recognition of the steps common to all testing programs: (1) selecting appropriate and usable tests, (2) giving the tests, (3) scoring the tests, (4) analyzing and interpreting tests results, (5) recording test results, and (6) using and applying test results.

Each step is discussed separately in the chapters which follow. The specific point to be made at this time is that to be complete and comprehensive the overall testing plan should provide for efficient carrying out of each of these steps. Allowing for some degree of flexibility, these items can be planned for in advance.

The fourth specification needs little elaboration. The program must, first of all, be practical in terms of limitations imposed by budget and personnel. A minimum program is better than no program at all. Secondly, the program must be definite. In stating objectives and defining the specific purposes to be served by it, definite terms should be used, employing meaningful descriptions of pupil behavior and characteristics identifiable in concrete terms. The plan for carrying out each testing procedure should be definite in assignment of duties, indication of time and place, and so forth. This will avoid confusion, duplication, ambiguity, and wasted effort.

For a test program to become effective it must be continuous. The last stated essential for adequate program planning refers to this factor of continuity. Sporadic testing efforts seldom contribute to effective guidance, nor is individualized instruction made possible by the results of a single testing program. The amount and direction of individual growth and development can be determined only by evidence which accumulates over a period of years. Hence, effective overall planning will reach far ahead of the

INTRODUCTION TO TESTING

general objective. A more detailed study of long-range planning in the community school discussed in Chapter 5. The designing of such a program will enter into long-range planning intended to introduce a program of testing and evaluation.

WHAT CONSTITUTES A MINIMUM TESTING PROGRAM?

In absorbing the costs of the testing program, the private or independent school may be able to increase the individual tuition charge by a small amount, thus passing the cost of special guidance service directly on to the parent. The public school, however, is not in a position to follow this practice and must, in fact, watch rather carefully each budget allocation. Because cost is of primary concern, the public school educator frequently is faced with a problem of this sort: "My testing costs this year must remain within a certain limited amount. What are the minimum testing requirements I should attempt to provide with this budget?" or "What are the maximum benefits I can provide with this amount?"

It is somewhat academic to speak of a testing program apart from the situation in which it is to be applied. Differences in local objectives, in individual and community needs, in number and quality of staff members all operate to make ineffective in one situation what proves adequate in another. The extent to which minimum needs are met also will be determined partly by the purpose of testing. This may be to assist with the business of individualized instruction, or it may be perhaps to check on the status of a particular school or school system when compared with regional or country-wide norms. However, assuming that the primary purpose of testing is to help the teacher achieve

a better understanding of the pupil and realizing that there are certain basic needs common to most educational situations, perhaps it is possible to deal generally with the relative importance of various parts of the testing program.

In connection with this topic, several public-school members of the Educational Records Bureau were asked to describe local testing programs and to indicate modifications that might occur if the various plans of testing were reduced to a bare minimum. Variations with regard to both current programs and concepts of minimum program were revealed in the replies. However, there were notable points of agreement:

1. All agreed that a minimum program would include testing of mental ability. Further, it was generally agreed that mental testing should be applied at more than one point in the educative process.

2. At least one type of achievement test, the reading test, was mentioned in all minimum program descriptions.

3. In those situations where current plans include a general program and a supplementary program (the latter being mostly individual and small group testing) there was general agreement that the supplementary program should be eliminated before the general program.

4. No description of a *minimum* program included interest testing or testing of personality.

The Bureau's experience in dealing with measurement and guidance programs over a number of years would tend to support the points of agreement in the informal survey just described. If it were necessary to limit the testing program to a single type of test, probably a mental ability test would be selected, preferably one designed to give a diagnostic picture of capacity. It is difficult to get accurate informa-

concerning individual capacity for learning by any other means. Teacher-made evaluation devices will, within the limits of local comparison, provide fairly adequate information about achievement. Observation of pupil behavior yields useful data in some of the other areas where information is needed. But it is virtually impossible for the teacher to determine accurately by observation whether unsatisfactory growth is due to lack of capacity or to one or more other factors. A fairly accurate description of individual mental ability is basic to individualized instruction. General ability testing, then, would constitute the barest minimum of objective testing.

A minimum program which employed only a test of general mental ability would need at least two provisions in order to be effective. First, it would have to be continuous to avoid mistaken decisions on the basis of a single test result. An intelligence test given in the early school months should be repeated at least near the transition from primary to intermediate grades and again somewhere near the time of change from elementary to secondary grades. Second, retesting should be planned on an individual basis whenever the initial result is considered unreliable. Ideally, the follow-up test should be an instrument designed for individual administration, such as the Stanford Binet Scale. However, if this procedure is not possible, a separate form of the group test or a different group test should be administered.

If the local situation allows use of only two objective tests, a reading test probably should be the second kind of test used. Since reading provides the medium through which much of the pupil's learning takes place, information concerning individual achievement in this tool subject is of considerable importance. Accurate

measurement of reading rate and comprehension can, in itself, without necessarily over-stated and well-standardized measuring instruments. Objective tests are available which will not only identify those pupils having reading difficulties but also assist in diagnosing the particular difficulty leading to unsatisfactory reading achievement. Tests are available also to evaluate generally the individual's readiness for reading instruction.

If one attempted to develop a broader and broader program based on these suggested minimum essentials, he would find that local differences began to make greater and greater impact upon adequacy of planning. Probably one would amplify the testing program for general ability and reading by adding other objective tests of growth and development in various subject matter areas. One would wish perhaps to add achievement tests in tool subjects other than reading before attempting measurement in broader subject matter areas where instructional objectives will vary from school to school.

Development and application of more comprehensive testing programs usually require that some centralization be given to the testing function and that assignment of responsibility be centered in some person or group of persons especially qualified in testing. With provision of adequate staff, the general ability and achievement testing program can be supplemented by instruments to provide information concerning interests, special abilities and aptitudes, and habits and traits of personal and social adjustment.

In considering an ultimate or "maximum" plan of testing which might result from continuous addition to minimum essentials, it may be of interest to examine the fall and spring testing programs sponsored by the Educational Records

INTRODUCTION TO TESTING

Bureau. They are shaped generally by a committee consisting of teachers and testing and guidance specialists drawn from member schools.

In brief outline form, this program is as follows:

1. *Academic aptitude.* Various group tests of general intelligence are recommended for use at each level from Grade 1 through Grade 12 in the fall. In instances where pupils' scores are much lower than had been expected, it is suggested that an individual test, another form of the test originally employed, or a different group test of mental ability be given. For measurement of academic aptitude of pupils with reading difficulties, nonverbal tests or tests with nonverbal sections are recommended.
2. *Reading.* Standardized reading tests are recommended for use at each level from Grade 1 through Grade 12, both in the fall and in the spring. It is recommended that further diagnostic testing be applied in instances where pupils show reading deficiencies. A reading readiness test is recommended for use in kindergarten and early Grade 1.
3. *Spelling.* Diagnostic spelling tests are recommended for use in Grades 4 through 12, both in the fall and in the spring.
4. *General achievement.* Tests are recommended for use in Grades 2 through 8, providing measures of achievement in tool subjects in lower grades, with tests in science, social studies, and literature added in intermediate and upper grades. These tests are recommended for use both in the fall and in the spring. Specific subject matter achievement tests are recommended for use in Grades 9 through 12, in the fall for placement of new pupils and in the spring for measurement of growth in subject matter knowledge.
5. *Vocational Interests.* The Bureau program provides inventories of vocational interests for use in Grades 9 through 12, either in the fall or in the spring.
6. *Special aptitude.* The testing of mechanical aptitude and mechanical comprehension is provided in the Bureau program.
7. *Diagnostic tests.* In addition to the diagnostic reading and spelling tests already mentioned, diagnostic mental ability tests yielding profiles of mental ability scores are made available. Also, diagnostic testing is provided for language and arithmetic skills and for study habits.
8. *Experimental program.* Usually some new test is employed in the experimental part of the program. Recently an inventory of youth problems was offered as an experimental instrument. Other types of tests used in this part of the program have included tests of primary mental abilities, tests of interests, survey of study habits, and new reading and achievement tests.

Participation in the Bureau program is not on an "all or none" basis. In other words, schools can select any part of the program which is desirable and practical in terms of local measurement needs.

This overall plan of testing illustrates a "maximum" testing program in the areas of academic aptitude, achievement, and interests. Not much could be added in terms of quantity or volume, although it will be noted that there is little mention of personality tests and that special ability tests are perhaps underemphasized. No doubt this lack of emphasis reflects the special interests and needs of the independent-school group, where non-college-preparatory pupils are in the minority and the need for tests of a "vocational" nature is at a minimum. Also, in the smaller independent-school classes it is easier perhaps to

HOW SHALL WE PLAN A TESTING PROGRAM?

employ with some effectiveness the techniques of teacher ratings, anecdotal records, and observations of pupil behavior in evaluating personal qualities. This practice may draw interest away from objective tests of the personality type. Actually, the members of the Committee on Tests and Measurements have been wary of the personality scales because of the difficulties mentioned earlier in establishing satisfactory validity for the instruments now available.

It should be pointed out again that although consideration of minimum and maximum programs may afford some structure within which local planning may operate, by far the best scheme in devising a plan of testing is to start first of all with local objectives and local needs and to orient the program toward these purposes. Thus, whether the program is maximum or minimum will be determined in terms of the following: "What are our first needs, those second in importance, and so on?" This question is

then answered in relationship to the specific situation.

SUMMARY

Certain items have been suggested as being characteristic of good test program planning. These include definiteness of purpose, coöperative planning, completeness of planning, thoroughness, practicality, and continuity. In providing for minimum testing requirements and then leading into a broader, more comprehensive program, one will usually begin with testing of general aptitude and testing of reading achievement, gradually developing upon this base broader achievement testing, testing of general abilities, and finally application of interest, character, and personality tests. The testing of the Educational Records Bureau is described as an illustration of the more comprehensive plans of measurement.

SUGGESTIONS FOR FURTHER READING

1. Erickson, Clifford E., *A Basic Text for Guidance Workers*, New York, Prentice-Hall, Inc., 1947, pp. 413-447.
2. Froehlich, Clifford P., and Benson, Arthur L., *Guidance Testing*, Chicago, Science Research Associates, 1948, pp. 5-10.
3. Micheels, William J., and Karnes, M. Ray, *Measuring Educational Achievement*, New York, McGraw-Hill Book Company, 1950, pp. 79-102.
4. *Planning a Testing Program*, Test Service Bulletin No. 55, Yonkers, World Book Company (undated), pp. 1-10.
5. Remmers, H. H., and Gage, N. L., *Educational Measurement and Evaluation*, New York, Harper & Brothers, 1943, pp. 19-37.
6. Ross, C. C., *Measurement in Today's Schools*, New York, Prentice-Hall, Inc., 1947, pp. 175-183.
7. Traxler, Arthur E., *Techniques of Guidance*, New York, Harper & Brothers, 1945, pp. 155-163.

How Can Tests Be Selected?

WIDESPREAD use of tests during recent years has been accompanied by an ever increasing volume of available testing instruments. During the half-century or so since the technique of objective testing had its inception, literally thousands of tests have been prepared and published. The *Fourth Mental Measurements Yearbook* lists 705 testing instruments, most of which were published fairly recently. One educational institution has catalogued some 1400 instruments making up its test library. As with any marketable product, one may assume that not all tests are equally good. Also, as with other consumer products, one may assume that there are certain qualities or characteristics which differentiate the better from the poorer instruments. The first task, then, in selecting tests for any stated purpose is to know just what to look for in choosing a satisfactory test.

WHAT ARE THE CHARACTERISTICS OF A GOOD TESTING INSTRUMENT?

If one examines current testing literature, he will find that four qualities are described frequently in dealing with test evaluation, that is, with an evaluation of the instrument itself.

These characteristics are (1) validity, (2) reliability, (3) objectivity, (4) usability.

VALIDITY

Validity is the degree to which a test measures what it is supposed to measure. This is viewed generally as the most important characteristic of a measuring instrument. Even though other qualities are possessed in a high degree, if the test lacks validity for the purpose intended, it must be discarded as unsuitable. For example, if it is desired to measure knowledge of geography facts, an instrument which contains many long paragraphs calling for a great deal of reading may not be considered valid for the purpose intended, since the score may reflect in large measure reading comprehension rather than mastery of geography facts. On the other hand, a test of this kind might be highly valid if the purpose were to measure ability to read materials in the field of geography instead of mastery of the essential facts. The *purpose* is the determining factor in validity.

One might assume that the name given to a test would tell what it measures. Thus, if it is desired to evaluate the mechanical ability of individuals in a particular group, one might simply consult test catalogues for listing of mechanical ability tests and expect that an in-

HOW CAN TESTS BE SELECTED?

strument chosen from the appropriate section would yield a measure of the factor which constitutes the title. Unfortunately, this supposition is not always true. One governmental agency, for example, found that a well-known mechanical ability test proved quite effective in differentiating better and poorer workers in a particular clerical job—in fact, it was much more valid for this purpose than any so-called clerical ability test which was tried. Only through systematically relating test content to curricular objectives or by statistical comparisons of test results with other criteria of what the test is supposed to measure can validity be demonstrated. The first method mentioned, that of comparing test content with stated objectives or stated purposes of testing and rendering some judgment regarding the degree of relationship, is generally known as curricular, or face, validity. This concept is important particularly for achievement testing. Face, or curricular, validity *can* be determined by the teacher or by the testing committee. If objectives are clearly stated and the purposes of testing are definitely identified, usually it can be determined, without technical background, whether or not the test under consideration is valid for the purpose intended.

The second method, that of statistical comparison, involves the procedure of computing correlations between test scores and other criteria of the characteristic to be measured expressed in quantitative terms. For example, statistical validity is sometimes investigated by correlating scores on a new test with scores on a similar test which has been used extensively and for which a significant degree of validity has been reported. Classroom marks and teacher ratings or judgments expressed quantitatively are examples of other criteria used frequently in

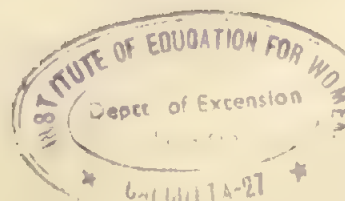
studying the statistical validity of an instrument.

Interpretation of a validity coefficient is dependent upon a number of factors, including reliability of the test, reliability of the criterion, homogeneity of the group studied, and other items best understood by the statistician. For this reason it is probably best done with the advice of those staff members as consultants who have appropriate technical background. However, the educator and the teacher not familiar with the technical aspects of validity can be trained at least to question the value of tests for which no validity data are offered or for which validity descriptions are obviously evasive and incomplete.

RELIABILITY

A synonym for reliability is consistency. A test which is reliable will yield approximately the same results upon repeated administration or when two closely comparable forms of the test are administered. If one should measure the length of a table with a cloth measuring tape several times, the results probably would agree rather closely. Some disagreement might occur as a result of fluctuations in the observation and alertness of the person using the tape. The cloth tape would be considered a reliable measuring instrument. The same table might be measured with an elastic tape and if the measurements were repeated notable differences in results might occur because of changes in length of the instrument itself. This would obviously be an unreliable instrument for measuring length. Reliability, then, refers to the extent of agreement one can expect in repeated trials of the test or other measuring instrument.

A clear-cut distinction should be made between the characteristics of validity and reliability. One refers to *appropriateness*; the other



INTRODUCTION TO TESTING

refers to *consistency or accuracy*. An instrument might possess a high degree of consistency or reliability while possessing very little validity with respect to a given testing objective. A high degree of reliability, then, in itself does not guarantee that a test is good. On the other hand, a test which possesses unsatisfactory reliability cannot be selected as a good test. One testing authority, C. C. Ross, summed up the situation by saying, "The ideal test tells the truth consistently."¹

When one examines test manuals or test reviews in order to determine the reliability of a particular test, he will find that this characteristic is reported in statistical terms, usually in the form of a reliability coefficient of correlation. This coefficient may have been determined in one of several ways. Among the methods employed in estimating test reliability are the following:

1. The Test-Retest Method. The test-retest procedure involves a second administration of the same test to the same group, usually with sufficient time between the two administrations to allow the pupils to forget most of the specific content of the test. Too long a period should not be allowed between the two test sessions, however, for learning will cause changes in the characteristic measured and the changes will be reflected in the results. Since it is a somewhat delicate task to strike a proper balance between these two complicating factors, the test-retest method is not used widely in reporting reliability coefficients.

2. Testing with Comparable Forms of a Test. The second method of studying reliability is somewhat similar to the test-retest method. It requires that closely comparable parallel forms

of a testing instrument have been constructed, attention having been given to the degree of similarity in such aspects as item validity, item difficulty, mental processes required for answering the items correctly, and sampling of subject matter. As with the first method, a time interval is usually provided after the first form is given, although this interval is generally shorter than when the test-retest method is used. After the second form is administered, the results of the two forms are correlated. What is known as "practice effect" is a somewhat complicating factor. Practice effect means that the experience of taking the first form of the test is likely to cause pupils to do better on the second form. If the effect were the same for all pupils, the correlation would not be affected, since a constant would simply be added to the scores on the second form; but it is known that some pupils profit from practice more than others.

A second difficulty issues from the assumption that the separate forms are completely comparable. One needs to know the extent of agreement in item validity and difficulty, and so forth, before he is able to interpret accurately the reliability coefficient resulting from the comparable-forms method of securing reliability. It is important for test users to have in mind the fact that reliability coefficients based on the administration of alternate forms of a test usually are somewhat lower than those resulting from other methods, particularly the Spearman-Brown method.

Notwithstanding the difficulties just mentioned, the alternate-form method is, from the standpoint of appraisal of the growth of pupils, perhaps the most defensible method of determining reliability. It is the most closely related to the way in which the test generally will be used in an actual situation. Cumulative records

¹ C. C. Ross, *Measurement in Today's Schools*, New York, Prentice-Hall, Inc., 1947, p. 83.

HOW CAN TESTS BE SELECTED?

for guidance purposes are likely to be based on the use of a series of comparable forms of a test over a period of years. Alternate-form reliability—that is, the correlation between successive forms—is therefore a matter of practical importance. The main reason it is used less frequently than the Spearman-Brown method described in the following section is that it is considerable bother to arrange for repeated use of the same test with the same pupils for the purpose of investigating reliability.

3. Spearman-Brown Method. Another method of determining test reliability is based upon a single administration of the test. The test is then split into chance halves, usually by scoring odd items and even items separately. By assuming equivalence of the two halves, one has in effect two closely comparable forms. The scores on the two halves are correlated and the coefficient is increased by application of a statistical formula devised independently by Spearman and Brown (hence the name of the method), yielding an estimate of the degree of relationship that would be obtained had each half of the test contained the same number of items as found in the whole test. Spearman-Brown reliabilities are reported quite frequently in test descriptions. Although the assumption that odd and even items yield equivalent halves may be difficult to defend, Spearman-Brown estimates have been found to agree rather closely with reliabilities obtained from the controlled test-retest method of estimate, where time is not an important factor in the administration of the tests. With highly speeded tests, the Spearman-Brown method nearly always overestimates reliability. Spearman-Brown reliabilities reported in test manuals for closely timed tests should seldom be accepted at their face value.

4. Kuder-Richardson Method. The Kuder-Richardson method is another means of estimating test reliability by application of a statistical formula. Coefficients derived through this method should be interpreted with considerable caution. Data drawn from a single administration of a test are employed. These are the mean of the test scores, the standard deviation of the test, and the number of items in the test. The procedure involves several assumptions which may not be valid with some tests. However, it is generally recognized that the technique tends to underestimate reliability. Hence, reliability data reported in terms of Kuder-Richardson coefficients are, generally speaking, conservative estimates of consistency, when compared with those obtained by the Spearman-Brown method.

What is acceptable reliability?

Interpretation of the reliability coefficient is again a task for the person having special preparation in the field of measurement. However, knowledge of the meaning of the term and of the techniques commonly employed in determining reliability will assist the untrained teacher or educator to make fairly adequate judgment regarding relative superiority of different test reliabilities. Generally speaking, minimum satisfactory test reliabilities are somewhat as follows: (a) For group prediction—that is, for estimating future group accomplishment—the reliability coefficient should be no lower than .60 and preferably should be at least .8. (b) For individual prediction—that is, the measurement of individual differences and prediction of future individual accomplishment—the reliability coefficient should not be below .9 and ideally should be .95 or above. However, many widely used tests do not measure up to this theoretical standard, even with regard to total scores,

INTRODUCTION TO TESTING

and extremely few meet this standard so far as scores on parts of the test are concerned.

The standard error of measurement for a particular test has occasionally been reported as an indication of its reliability. The Cooperative Test Division of the Educational Testing Service,² for example, reports the standard error of measurement for each of the tests in the Cooperative series. This index of accuracy or consistency is different from the reliability coefficient, both in interpretation and in derivation. Derivation of the standard error probably is best left to the statistician. With regard to interpretation, this index of accuracy provides clues as to the upper and lower limits within which a given score may be expected to fluctuate. To illustrate: The standard error of measurement for a Scaled Score of 50 on the Cooperative Intermediate Algebra Test, Form Z, is 3 Scaled Score units.³ That is, the true score probably lies somewhere between 3 points up from 50, or 53, and 3 points lower than 50, or 47. Actually, the statistical interpretation would be that the chances are somewhat better than two to one that the theoretical "true" score lies within this range. This is an indication of the accuracy or reliability of the instrument yielding the score. The standard error of measurement is a convenient means for showing that the reliability of measurement is not the same for all scores. On some of the scoring keys for the Cooperative tests the standard error of measurement is shown not only for a Scaled Score of 50 but also for a Scaled Score of 70 (two S.D.'s above the mean for the standard group) and for scores at other points in the distribution.

Since it is expressed in terms of the amount of fluctuation to be expected in a score obtained

by a particular individual, this system of reporting reliability may actually be more meaningful than reliability coefficients, and it would perhaps be desirable for test publishers to use it more frequently. However, the person not trained in statistics would have difficulty comparing different tests with respect to reliability if the index of consistency were stated for one as a reliability coefficient and for the other in terms of standard error of measurement. Examination of test manuals and reviews reveals that coefficients of reliability are used almost always in describing the accuracy or consistency of tests. Hence, the neophyte will do well to concentrate on some understanding of the reliability coefficient so as to be able to examine it more critically.

OBJECTIVITY

Objectivity of a test is the degree to which it can be scored with a minimum of individual judgment as to the correctness or incorrectness of responses to the test items.

The degree of objectivity of any standardized test can usually be determined by examining scoring procedures. Such examination reveals that some so-called objective tests provide considerable flexibility in stating acceptable responses. In some instances subjective judgment is required to distinguish correct responses—for example, on parts of certain tests in which the pupil is required to write out statements completing sentences or answering questions. This kind of answer obviously must be evaluated somewhat differently from a response made by selecting one of a group of suggested answers of which only one answer is correct.

Objectivity in a test is usually accomplished through careful item trial. Item analysis based on experimental tryout of the test and analysis

² Addresses of publishers are given in the Appendix.

³ Scaled Scores are described on pp. 53-54.

HOW CAN TESTS BE SELECTED?

and criticism by subject matter experts usually provide adequate information for phrasing items and item choices specifically so that all but the intended correct responses are for one reason or another wrong.

It may be argued that no test can be completely objective and that the measurement process requires subjective judgment at many points, especially in the choice of item content. Actually, one should not place too high a premium on objectivity, certainly not to the extent that validity or the effective measurement of educational objectives is sacrificed. Nonetheless, the objective testing movement has aimed at the elimination of needless subjectivity in measurement and evaluation. The advent of machine scoring has emphasized the need for attention to this characteristic in evaluating and selecting testing instruments.

USABILITY

A number of practical considerations tend to differentiate testing instruments, and to these the teacher or administrator will attach a good deal of importance, particularly in the public-school situation. They may be grouped together as those characteristics which make one test more usable than another. Among them are the following:

1. **Cost.** The cost of a particular test is not covered completely in the listed charge for the test booklet. Actually, the cost factor overlaps several of the other practical characteristics which will be mentioned. For example, some tests require specially trained persons to administer them. The hiring of such experts adds to testing costs. Scoring time and required scoring procedure affect the cost of the program. Whether or not separate answer sheets are provided with a particular test is another factor

affecting costs. When separate answer sheets are provided, booklets can be used over and over again so that initial cost can be written off over a period of time. Separate answer sheets are relatively inexpensive. This advantage is partially offset by the clerical cost of inspecting the test booklets after each use to make sure that pupils have not written in them. All of these items warrant careful consideration in selecting tests. It would be unfortunate if adequacy of measurement were sacrificed in favor of cost, but, other things being equal, one will aim for selection of that test which provides adequate measurement at least expense.

2. **Ease of Administration.** Ease of administration is a practical consideration which is important from the standpoint not only of cost but also of program planning. Some tests require about the same amount of mental gymnastics on the part of the examiner as on the part of the examinee. This situation is approached when the test is broken into several parts, each calling for carefully timed short working intervals, frequently with the examiner participating in the attack on each separate test item. Tests range in ease of administration to the other extreme, where the instrument is virtually self-administering, requiring little supervision except encouragement on the part of the teacher or examiner. Other things being equal, the test which is difficult to administer will be passed by in favor of a simpler test.

3. **Ease of Scoring.** In addition to figuring in the costs of the program, ease of scoring also is related to program planning. Can the test be scored locally, or will the services of a scoring agency be required? If it is scored locally, how much training will be necessary to handle the mechanics of the procedure? Is special machine processing required? If grade equiva-

INTRODUCTION TO TESTING

lents or standard scores are employed, are these conveniently identified?

Some tests, such as the Kuder Preference Record, have comparatively simple scoring arrangements so that the pupils themselves are able, under supervision, to score their own tests, provided the scoring is checked. Other instruments, such as the Minnesota Multiphasic Personality Inventory and the Strong Vocational Interest Blank, involve plus and minus weightings with several scoring stencils and usually require machine processing.

Surely one would not wish to sacrifice validity or reliability in favor of scoring arrangement, but, other things being equal, one would select that test which is easiest to score.

4. Printing and Test Format. Printing and format should receive rather careful consideration in selection of a testing instrument. Actually, limitations and defects with regard to typography and make-up may bring irrelevant factors into the test situation, such as visual acuity, resistance to fatigue and monotony, and other kinds of distractions. Quality of paper, legibility of type, arrangement of item stems and other printed matter in relationship to responses and relevant questions, use of pictures and illustrations, accuracy and consistency of directions supplied in the test booklet—all of these are important items which tend to differentiate good tests from the less desirable instruments. For years the Test Selection Committee of the Educational Records Bureau consistently refused to recommend an otherwise promising test of mental ability because the format was regarded as particularly bad.

5. Adequacy of Norms. The usefulness of any test is conditioned in large measure by the kind of normative data supplied to the test user. First of all, one needs to know if the

types of comparison provided are those desired. For example, a test would be relatively useless in the school situation if only general adult norms were provided. The representativeness of the population upon which the norms are based is also a matter of concern. Grade norms based on results contributed by pupils in one geographical area, for example, may provide a poor basis of comparison for pupils in another area.

Some test publishers provide separate norms for separate parts of a test in addition to total score comparisons. Norms may also be set up for age groups, grade groups, sex groups, occupational groups, or by region. The kind of groups upon which they are based must be evaluated in relation to the purposes of testing and kinds of comparisons desired. Thus, adequacy of norms is of definite importance as a practical consideration.

WHO SHOULD SELECT THE TESTS?

In describing the characteristics which differentiate available tests, we have suggested that certain aspects of analysis and interpretation are probably best left in the hands of a testing specialist. Hence, the director of guidance, the director of testing, or some member of the educational staff having special preparation in the technical aspects of measurement may figure largely in the selection of evaluative instruments after the purposes of testing have been set up. In the absence of such specialized personnel, frequently an outside consultant is called in to assist with this step in the testing program.

Supplementing the service of the test specialist, however, is the necessary contribution

HOW CAN TESTS BE SELECTED?

to be made by the subject matter expert. Thus, the curriculum committee may serve an important function in test selection by determining the appropriateness of particular instruments in relationship to the stated overall educational objectives of the school. Each department, or each individual teacher, should also contribute to test selection in determining the soundness and adequacy of test content with regard to specific educational objectives.

Selection of tests, then, may well be a coordinated, group enterprise with each of several groups making a definite contribution. It is not difficult to defend the point of view that any decision involving subjective elements is best made by competent group judgment rather than by any one person.

WHERE CAN ONE GET INFORMATION REGARDING TESTS?

In attempting to evaluate various testing instruments, one needs to know where he can go in order to obtain accurate information regarding test validity, reliability, usability, and so forth. The following sources are suggested:

1. THE MENTAL MEASUREMENTS YEARBOOKS.

Under the editorship of Oscar Buros, the *Mental Measurements Yearbooks* have been prepared to present listings of standardized tests and critical evaluations of each of the instruments listed. These yearbooks have been published in 1938, 1940, and 1949. They form the most complete and comprehensive source of information now available concerning tests, although other sources, such as Hildreth's bibliography,⁴ provide a more nearly complete

⁴ Gertrude H. Hildreth, *A Bibliography of Mental Tests and Rating Scales*, New York, The Psychological Corporation, 1933, 1939. (Now out of print. Lists more than 4000

listing of published tests. However, most of the commonly used instruments will be found in the listings of Buros' three volumes. Although the test reviews vary considerably in quality and in objectivity, they provide valuable information concerning a wide variety of tests.

2. TEST MANUALS

The manuals usually provided by test publishers with specimen copies of tests ordinarily contain information concerning test reliability and test validity. Quite frequently, sections of a manual deal with descriptions of norms, suggestions concerning use and interpretations of test results, and other items which are helpful in evaluating tests. However, it must be kept in mind that the test author intends to present his instrument in as favorable light as possible. One may not get complete information regarding limitations of the instrument from the test manual.

3. CATALOGUES OF TEST PUBLISHERS

Listings of tests under various classifications can be obtained from test catalogues. Usually a small descriptive paragraph indicates the purpose of the test, the level for which it is intended, the kinds of norms that are available, the administration time, the availability of separate answer sheets and other scoring information, and, occasionally, information regarding test reliability. Again, it must be remembered that test catalogues are intended to publicize those instruments handled by a particular commercial organization; hence, little information can be expected regarding weaknesses of the instruments. Listings and descrip-

titles. A supplement issued in 1945 lists over 1000 additional titles.)

INTRODUCTION TO TESTING

tions provided in test catalogues, however, provide a good starting point for identifying instruments which may satisfy practical requirements. Evaluation of other characteristics can then be carried out by reference to other sources, such as the *Mental Measurements Yearbooks*. Test catalogues can be obtained simply by writing the test publishers, most of whom are listed in the Appendix.

4. TESTING AGENCIES AND ORGANIZATIONS

A number of agencies offer advisory service to schools planning testing programs. Among these are the United States Office of Education, the various state departments of education, state universities, and other public and private institutions of higher learning. Non-profit organizations, such as the Educational Records Bureau and the Educational Testing Service, offer assistance in setting up testing programs. The major commercial test publishers also maintain professional staff members to deal with requests for assistance in setting up plans of testing. Ordinarily, information is given by these organizations and agencies through correspondence, although occasionally a visiting consultant can be provided for special programs.

5. PROFESSIONAL LITERATURE

The various professional periodicals and testing publications provide valuable information regarding test reliabilities and validities, descriptions of new instruments, special uses of tests, and other kinds of research information.

It is something of a task to classify and digest professional literature in order to obtain the exact kind of information desired. The publication *Psychological Abstracts* is helpful in this connection; it classifies and abstracts current research articles and publications, offering material assistance in the task of locating source materials for a specific purpose or topic. A listing of the important periodicals which provide test information would include, among others, *Educational and Psychological Measurement*, the *Review of Educational Research*, the *Journal of Educational Research*, *Occupations*, and *Educational Records Bulletins*.

SUMMARY

Tests are selected usually on the basis of four major characteristics: (1) validity, (2) reliability, (3) objectivity, and (4) usability. Validity refers to the trueness of the test or its usefulness for a particular purpose; reliability concerns consistency of results; objectivity is that element of test make-up which tends to eliminate subjective judgment in scoring; and usability refers to practical items such as cost, ease of administration and scoring, printing and format, and adequacy of norms. Test selection may well be a coöperative enterprise, involving contributions from the test expert, the subject matter expert, and the school administrator. Important sources of information regarding tests include *Buros' Mental Measurements Yearbooks*, test manuals, test catalogues, testing agencies, and professional literature.

HOW CAN TESTS BE SELECTED?

SUGGESTIONS FOR FURTHER READING

1. Buros, Oscar K., *The 1940 Mental Measurements Yearbook*, Highland Park, New Jersey, *The Mental Measurements Yearbook*, 1941.
2. Buros, Oscar K., *The Third Mental Measurements Yearbook*, New Brunswick, New Jersey, Rutgers University Press, 1949.
3. Cronbach, Lee J., *Essentials of Psychological Testing*, New York, Harper & Brothers, 1950, pp. 43-83, 270-302.
4. Darley, John G., *Testing and Counseling in the High School Guidance Program*, Chicago, Science Research Associates, 1943, pp. 88-128.
5. Davis, Frederick B., "Two New Measures of Reading Ability," *Journal of Educational Psychology*, 33:365-372 (May, 1942).
6. Froehlich, Clifford P., and Benson, Arthur L., *Guidance Testing*, Chicago, Science Research Associates, 1948, pp. 11-46.
7. Greene, Edward B., *Measurements of Human Behavior*, New York, The Odyssey Press, Inc., 1941, pp. 97-108, 601-637.
8. Greene, Harry A., Jorgensen, Albert N., and Gerberich, J. Raymond, *Measurement and Evaluation in the Secondary School*, New York, Longmans, Green and Company, 1944, pp. 52-73, 109-114.
9. Lindquist, E. F. (ed.), *Educational Measurement*, Washington, American Council on Education, 1950, pp. 417-454, 560-694.
10. Micheels, W. J., and Karnes, M. R., *Measuring Educational Achievement*, New York, McGraw-Hill Book Company, 1950, pp. 103-125.
11. Paterson, Donald G., Schneider, Gwendolyn G., and Williamson, E. G., *Student Guidance Techniques*, New York, McGraw-Hill Book Company, 1938, pp. 52-256.
12. Ross, C. C., *Measurement in Today's Schools*, New York, Prentice-Hall, Inc., 2nd ed., 1947, pp. 183-193.
13. Traxler, Arthur E., *Techniques of Guidance*, New York, Harper & Brothers, 1945, chaps. 4, 5, and 6.
14. Wood, Ben D., and Haefner, Ralph, *Measuring and Guiding Individual Growth*, New York, Silver Burdett Company, 1948, pp. 229-260.

How Should Tests Be Given?

IT CANNOT be stated too strongly that correct administration of the tests is basic to any testing program. Tests administered carelessly or in such a manner as either to give pupils an unfair advantage or to put them at a disadvantage will yield invalid results no matter how good the instruments may be. With this warning, it should be pointed out that any teacher who will take the pains to prepare for the task can learn to administer group objective tests in a highly professional way and with fairly dependable results. This chapter will discuss preparatory activities which can be planned for teachers, pupils, and other persons having testing responsibility.

There are two aspects of preliminary planning for actual administration of the tests. One is the overall schedule dealing with assignment of examiners to class groups, necessary modifications in the usual school routine of classes, distribution of supplies to examiners, and so forth. The second is the work of each examiner in actually giving the tests.

In overall planning, responsibility is usually either accepted by the principal or another administrator or assigned to some one member of the faculty who will serve as coördinator of the program. This person will be responsible for preparation of an examination schedule

dealing with details of location, timing, assignment of examiners and proctors, and other items which serve to coördinate the plan of testing. The effectiveness with which this work is carried out is one of the most important factors in the success of the testing program.

Schools sometimes ask whether the tests should be given in the regular class periods or according to a special testing schedule. The answer to this question will vary with the circumstances. If only two or three different tests are to be given and if the time limits are relatively brief, it may be possible to fit the tests into the regular classroom schedule instead of interrupting the routine with a special arrangement. On the other hand, if the program is a very comprehensive one, if the school is large, or if the time limits for the tests are somewhat longer than those for the class periods, a special testing schedule probably is to be preferred.

If a special schedule is followed, usually it is most convenient to extend the time over two or three days. During this period some classes will necessarily be dismissed, although, through careful planning, it is generally possible to have the classes not directly involved in the testing at any given hour meet on a regular basis. This procedure is probably desirable

HOW SHOULD TESTS BE GIVEN?

from the standpoint of maintenance of morale and encouragement of the pupils to take the tests "in stride." When a series of tests is to be administered within a relatively short period, there is some danger that pupils may become overtired if too many tests are given in one day. The ordinary academic aptitude and reading tests require about an hour of testing time each. Achievement test batteries, however, may require as much as four hours of work, perhaps too much to expect of most elementary-school pupils in one day. Probably the best way to administer such a series of tests is to give it in two periods of about equal length on successive mornings. If necessary, the test battery can usually be divided into still smaller units. If possible, the schedule should be so planned that no pupil is required to take more than four tests in a single day.

The use of separate answer sheets with objective tests complicates further the task of overall planning. Frequently, for purposes of economy, only enough test booklets are purchased to provide for the largest group to be tested at any one time, with the addition of sufficient answer sheets to account for all pupils. If the same test is to be administered to several different groups, the schedule must be staggered so that test booklets will be available as each group is tested. Time must also be allowed for inspection of the booklets after each testing to make sure that no answers or other marks have been entered in the pages of the booklets.

Probably the coordinator of the program will wish to prepare a mimeographed sheet outlining specifically all items of the overall test schedule. This sheet can then be distributed to the teachers who will serve as examiners. An illustrative sheet of this sort, giving general directions for a school testing program, is presented in one

of the references suggested at the end of this chapter.¹

The second aspect of preparation for giving the tests also needs emphasis. Ordinarily, responsibility as an examiner is assigned to the classroom teacher, who may or may not have experience in giving objective group tests. There are important general principles of test administration, mastery of which will enable the average teacher to approach the examination situation with confidence. These principles are intended to provide a test situation which will (1) call forth the pupils' best effort and (2) duplicate as nearly as possible the prescribed conditions for each test.

With regard to the first item, it is well known that scores obtained on any test are influenced by many factors other than the knowledge or understanding possessed by the pupil. Favorable testing conditions help to minimize fluctuations due to extraneous circumstances and to provide more accurate measures of the function being tested. Location of the testing room to avoid noise and distraction, the size of the room, lighting and ventilation, the arrangement of the chairs and tables or desks all are details which should be considered in order to make testing conditions as favorable as possible. The doors to the room should be marked to indicate that testing is going on. A *TESTING: DO NOT DISTURB* sign may well be used for such purpose.

In addition to these extraneous factors, favorable conditions also imply that proper motivation has been secured. The pupils themselves should be told enough about the purposes of measurement so that they realize the tests are given to help them, and that they should try to do their best.

One should minimize testing as the applica-

¹ See reference 9, pp. 159-161.

INTRODUCTION TO TESTING

tion of an external standard and should emphasize the guidance value of the results. Actually, the attitude with which the test is approached may have considerably more effect on the reliability of results than do heating, lighting, ventilation, or other outside factors. The pupil will wonder why the test is given, what will be done with the results, how his status will be affected by his test performance. Whether or not his feelings will result in carelessness, apprehensiveness, or even emotional block will depend upon the effectiveness with which the examiner establishes coöperation and uniform motivation and sets the group at ease.

The second general condition toward which principles of test administration are aimed refers to familiarity with test directions and the ability to follow them. The first step in this regard is to become completely familiar with the manual which is provided for each test and with the general format of the test itself. Probably the best way of preparing to give any test is for the examiner himself to take the test. By doing so, he can anticipate difficulties which may arise and foresee possible questions. At least the examiner should rehearse to the extent of reading aloud all directions and noting carefully the timing before he enters the examination room.

One of the advantages of having the examiner take the test before he gives it is that he then understands fully the timing of the test. Whenever a specific time allowance is given, it must be observed meticulously. If less than the recommended time is used and pupils have not had a chance to complete as much of the test as they are capable of completing, the results will not be comparable to those already on hand and the percentile norms will not be applicable. Likewise, if more time is allowed than the instructions call for, it will not be possible to interpret

the results of the examination according to norms previously established. The examiner should be equipped with an accurate timepiece. An ordinary watch with a second hand is satisfactory if timing entries are made on a slip of paper and the examiner does not depend upon memory. For a test with part timing, it is wise to make up a time sheet such as the following:

Test: Cooperative English

	<i>Part</i>	<i>Time</i>	<i>Start</i>	<i>Stop</i>
C	Vocabulary	15 minutes		
	Reading	25 minutes		
A	Grammatical Usage	15 minutes		
	Punc. and Capital.	15 minutes		
	Spelling	10 minutes		
B	Sent. Structure and			
	Style	15 minutes		
	Active Vocabulary	10 minutes		
	Organization	15 minutes		

When the examiner gives the word to "start," his eyes are on his watch and he immediately writes down the time in the appropriate column. The allowed time can be added to this immediately in order to make an entry of the exact time when the signal to "stop" is to be given.

In no instance should the examiner attempt to recite the test directions from memory. Omission of one word or phrase may change the test situation definitely from that prescribed in the manual. In giving the test, printed directions should be read verbatim from the manual. In going through the directions ahead of time, it is helpful to mark those places where special emphasis should be given. Directions should be read clearly and in a voice loud enough to be heard throughout the testing room.

The examiner can expect various types of questions to arise during the examination. The only kind of information he is allowed to give is in explaining the directions. He may not give

HOW SHOULD TESTS BE GIVEN?

any help in the solution of problems in the test. Again, any deviation from this rule will render the test situation quite different from that upon which the test has been standardized.

With large groups it is essential that testing assistants or proctors be assigned to help with answering questions, to assist in getting the pupils seated properly, to see that test supplies are distributed with dispatch, and to handle such details as supplying examinees with new pencils when leads are broken. The proctors should be careful to distract the pupils as little as possible. No unnecessary conversation between proctors or between proctor and pupil should be allowed, and the proctor should be instructed to move about no more than is necessary.

The matter of pupil preparation plays an important part in the standardized test situation. Because all succeeding steps of ranking pupils and classes, undertaking curriculum revisions, and proceeding with individual and group guidance depend on the accuracy of the test results as a picture of the pupil's status at the time of testing, it is particularly important that both teacher and pupil take the tests "in stride." Special coaching on the subject matter of a test, or other specific preparation resulting from overanxiety on the part of the teacher, can entirely invalidate the results of the soundest testing program.

Ideally, all pupils tested should be at about the same point of familiarity with the type of testing used. Pupils who have not taken objective tests previously can be told something about the form and general purposes of the tests. Such preparation is different from coaching and is necessary, because, otherwise, pupils who had taken objective tests previously would have an unfair advantage. Even the youngest pupils

tested should understand that some parts will be too difficult and some quite easy and that they must not worry if they are not able to answer all questions. Pupils may also be told what the general form of the questions will be—multiple-choice, matching, completion, or the like—and how much they will be penalized for wrong answers. A sheet of instructions to the pupil, covering the general purposes of testing, may be obtained from the Educational Records Bureau at a small charge. Some such statement as this prepared for the school's own situation may be distributed for pupils to read at the time the testing schedule is announced.

Some tests provide separate practice exercises to familiarize pupils with the form of the test before the testing proper begins. Unless the test provides for their use during the regular testing period, the practice tests may be distributed in advance, and pupils should be given an opportunity to obtain help if they have any difficulty. The examiner can inspect the completed practice tests in advance of the time set for the examination to assure himself that every pupil has understood the directions.

When tests are to be machine-scored, special instructions must be given to the pupils. These are usually covered in test manuals but are of sufficient importance to warrant separate emphasis. Machine scoring depends upon use of a special, soft lead with high graphite content. This pencil mark, made between pairs of dotted lines on the separate answer sheet, carries an electric current and serves to close a sensing unit circuit in the machine. It is essential not only that special pencils be used but also that examinees make *heavy, black, glossy* marks on the answer sheet. Stray marks and dots from doodling or figuring on the separate answer sheets, poor erasures, light or incomplete marks

INTRODUCTION TO TESTING

will affect the reliability of results produced by the machine. To prepare the pupils for use of machine-scored separate answer sheets, special practice tests have been prepared covering the various clerical aspects of marking responses. As with other practice tests, these can be distributed in advance of testing time in order to clear questions with regard to this part of the testing procedure. Even with this thorough preparation, it is advisable to spend a few minutes in rechecking the answer sheets following the close of the testing session. After the booklets have been collected, but before the sheet is handed in, each pupil can go over his marks in order to make them heavier and darker, can clean up stray marks and dots, and can make all attempted erasures thorough and complete. It is unfortunate indeed when a pupil fails to receive credit for a correct response owing to poor marking of the answer sheet.

To summarize: The routine of testing may seem complicated at the beginning. As with

any procedure involving the activities of a number of persons, however, effective preliminary planning will provide order and system for the whole process. Part of this planning involves overall coordination of the program, and part involves practice and preparation in the technical aspects of test administration. The latter phase usually requires some in-service training with regard to procedures which will yield dependable results as well as training concerning pitfalls which may tend to make the test scores meaningless. Elements pertinent to the test environment, such as arrangement of seating, illumination, freedom from distraction, and so forth, as well as proper motivation and attitude on the part of the pupil, need careful attention. Every effort should be made to duplicate as nearly as possible the standardized testing conditions which are described in the test manual. It must be emphasized that the usefulness of test results depends largely upon the care and accuracy with which the tests are administered.

SUGGESTIONS FOR FURTHER READING

1. Bingham, Walter V., *Aptitudes and Aptitude Testing*, New York, Harper & Brothers, 1937, pp. 224-237.
2. Cronbach, Lee J., *Essentials of Psychological Testing*, New York, Harper & Brothers, 1950, pp. 584-597.
3. Froehlich, Clifford P., and Benson, Arthur L., *Guidance Testing*, Chicago, Science Research Associates, 1948, pp. 47-49.
4. Lindquist, E. F. (ed.), *Educational Measurement*, Washington, American Council on Education, 1950, pp. 329-365.
5. Remmers, H. H., and Gage, N. L., *Educational Measurement and Evaluation*, New York, Harper & Brothers, 1943, pp. 486-509.
6. Ross, C. C., *Measurement in Today's Schools*, New York, Prentice-Hall, Inc., 2nd ed., 1947, pp. 193-199.
7. Super, Donald E., *Appraising Vocational Fitness*, New York, Harper & Brothers, 1949, pp. 72-85.

HOW SHOULD TESTS BE GIVEN?

8. Terman, Lewis M., and Merrill, Maude A., *Measuring Intelligence*, Boston, Houghton Mifflin Company, 1937, pp. 52-71.
9. Traxler, Arthur E., *Techniques of Guidance*, New York, Harper & Brothers, 1945, pp. 155-163.
10. The separate manuals issued by test publishers contain detailed directions for the administration of tests.

How Should Tests Be Scored?

OBJECTIVE tests have been said to yield the same results no matter who scores them. The accuracy of this statement depends, of course, upon freedom from error in securing and recording the scores. Errors may be due to carelessness on the part of the scorer or perhaps to lack of familiarity with, or to misunderstanding of, scoring instructions for the particular test. It is obvious that this is another point in the testing program where results may become unreliable and misleading. No matter how well tests are selected, no matter what the degree of professional skill applied in administering the tests, lack of planning and preparation for the scoring task may invalidate the entire program. Although scoring of objective tests is essentially a clerical task, scoring directions must be clearly understood, and the specified procedures must be followed precisely. This chapter will discuss general scoring problems and suggest preparatory procedures for this part of the testing program.

At the outset, it is necessary to decide whether the tests will be scored at the school or by a central agency. If the latter procedure is followed, the school is relieved of responsibility in planning and coordinating the scoring procedures. Although some large school systems are able to maintain test scoring departments, either staffed with scoring clerks or equipped with

scoring machines, probably most public schools will depend upon teachers to do the clerical work if the tests are scored locally. The advantages of teacher scoring and those resulting from scoring by a service agency may well be compared at this point.

Among the favorable aspects of having scoring performed by a central agency which is especially equipped to perform such service are the following:

1. If the tests are scored by an agency, the teacher has one less demand upon his busy time and can engage in more productive activities on a higher professional level.
2. The school is relieved of the supervisory responsibilities necessary in training scorers, checking accuracy of results, and so forth, which are involved in local scoring.
3. Usually there is less chance of error if tests are scored by a specialized agency.
4. The school will usually participate in and benefit from comparative analyses and other statistical research data provided by the central agency for commonly used tests.

On the other hand, certain benefits will result from teacher scoring of tests. Among these are the following:

1. The results may be available at an earlier time.

HOW SHOULD TESTS BE SCORED?

2. The teacher may obtain better understanding of the meaning of the results by actually scoring the papers and becoming familiar with the method by which the score is obtained.

3. Scoring by teachers provides an opportunity to participate directly in the testing program and may assist in establishing it as a group endeavor. This can be listed as an advantage only if the teacher is freed from other responsibilities while scoring is carried out and consequently is not resentful of additional work.

If tests are to be scored locally by the teachers, it is advisable to devote a period of group instruction to mastery of scoring directions. As indicated in an earlier chapter, the complexity of scoring procedures varies considerably from test to test. In all instances it is essential that the directions for scoring provided in the test manual be followed exactly as stated. Only those responses allowed in the scoring key, or as qualified by the test authors, can be credited as correct answers. The scorer cannot apply his own interpretation here. Some teachers may be inclined to penalize pupils for incorrect grammar, spelling, or punctuation in items where the answers are to be written in. Again, standardized instructions must be followed, and these usually state that no penalty is to be made for mistakes in spelling and grammar in tests other than those designed to measure spelling and grammar skills. Application of the scoring formula, if one is applied, use of norm tables, computation of medians or averages all are items which need emphasis in an instruction period in order to insure that all teachers follow a standardized procedure.

It is essential that a check on accuracy of scoring be made by some person other than the initial scorer. For this function, teachers may be asked to exchange papers and check on each

other's work. It may be agreed to rescore every other paper or perhaps every third paper, but if frequent errors are discovered, each paper in the group should be rescored.

Some schools carry out local scoring by carefully planned group procedure whereby all scorers assemble in one room and each is assigned a certain task in the scoring process, such as scoring certain subtests, transferring scores from test pages to the front of the booklet, or perhaps copying percentiles or grade equivalents corresponding to raw scores. The booklets are passed along from one scorer to another until all the processes are completed. Such a procedure requires either some experience with scoring the tests or else preliminary time analysis of the various elements of scoring.

If the tests are to be scored locally by the teachers, it may be advisable to dismiss school for a half-day, or perhaps two half-day sessions, depending upon the length of the scoring task, in order to complete processing of the tests. This procedure will enable the teacher to concentrate upon the task at hand and will tend to eliminate "fatigue" errors which commonly result in activities performed after a regular day's work.

The scoring procedures followed by the Educational Records Bureau in processing the thousands of tests which are handled in its offices each year take into account all the requirements of accurate scoring. While these are, no doubt, highly specialized in comparison with the usual local situation, they illustrate the operations that are involved in securing accurate results. The general procedures are as follows:

When test booklets are received, they are carefully counted and classified. Birth dates are checked and chronological ages computed. The school is notified of any discrepancies in statements about age or grade so that they can be

INTRODUCTION TO TESTING

corrected before a report is made. At the same time the Bureau checks to see that enough information is given to insure accurate classification of tests. This is particularly important in connection with achievement tests in mathematics and foreign languages. If the norms are to be meaningful, it is important to know, for example, whether a tenth-grade French class has been studying French for a semester and a half or whether French was begun in the ninth grade. In some situations, advanced algebra may be preceded by both an intermediate and an elementary course; in others only one year of study may form the background for what is called the advanced course. Therefore, each school is asked to indicate, for each class, the name of the course and the year of work which the course represents in terms of the school's own curriculum. Specific information about the amount of study completed in terms of years, number of periods per week, and length of period is also requested. Although this task may seem irksome, the school filling out the blank furnished for this purpose will realize that it is obviously in a better position to estimate the amount of study than an outside agency could possibly be.

The school is also asked to indicate its preference for the form in which the report is to be prepared. For example, large schools may prefer to have scores on certain tests listed by sections rather than by complete grades. Remedial cases or "special" pupils are identified so that their papers will be excluded from distributions of class results. Test results can be reported in the way which will be most helpful to the school if the school specifies what form of report is desired.

Figure 1 shows a typical classification slip. This slip has been filled out to attach to a group of forty-seven Cooperative Elementary Algebra

Tests, Form Z, received from the Brownville, New York public schools. The test was given to a ninth-grade class in beginning algebra. The processing operations performed by the Bureau staff are listed on the classification sheet. There is space opposite each item for the initials of the person who performs each task, and for timing entries. These slips are not detached from the test booklets until they are sent back to the school after the scoring has been completed and the report made.

The tests are now sent to the scoring department. Hand scoring of an objective test when the answers are entered in the test booklet consists of comparing the responses on the test with a key furnished by the publisher. All correct answers are marked in the margin with a short horizontal line and these marks are put in a vertical column. The wrong answers are marked only if the scoring formula calls for the subtraction of some proportion of the incorrect responses from the number of right answers. If wrong answers are marked, they are shown by X's, which are put in a vertical column to the right of the horizontal lines marking correct answers. All original scoring is done in red pencil. If the scorer makes an error and discovers it himself, he changes the wrong mark by putting a wavy line through it and writing the correct mark at one side.

Scorers are required to adhere strictly to the printed keys. If an error in the key is discovered, the error is called to the attention of the scoring supervisor, who checks the answer, changes all keys, and has all previously scored papers changed to agree with the corrected key. Such an error is, of course, called to the attention of the test publisher. The supervisor also advises the scorer in the case of ambiguous responses, illegibility, and so on. Ambiguity in an-

NO. A NO. PAPERS 47 GRADE 9
 SCHOOL BROWNVILLE HIGH YR. OF STUDY 1 YEAR
 LOCAL COURSE _____
 SUBJECT ELEM. ALGEBRA MR. CHRISTENSEN
 TEST Coop.
 FORM Z
 CLASSIFY MB CHECK LC HOLD, DISC. _____ ABSENTEES _____

PART	SCORE			REScore				COUNT			RECOUNT			
	INIT.	IN	OUT	INIT.	IN	OUT	ERRORS	INIT.	IN	OUT	INIT.	IN	OUT	ERRORS
1	FP	9:10	9:40	CT	1:05	1:11	9/10				CT	1:11	1:21	0
2														
3														
4														
5														
6														
7														
8														
9														
10														

OPERATION				CHECK			OPERATION				CHECK		
	INIT.	IN	OUT	INIT.	IN	OUT		INIT.	IN	OUT	INIT.	IN	OUT
Chron. Age							Alphabetize	BM	2:23	2:25	AP	3:09	3:11
Scan							Match						
Mark							Profile						
Transfer							Plot						
Add							Connect						
Sc. S. Parts							Class Prof.						
Sc. S. Total							Cl. Md. Sheet						
Recheck							Cl. Anal. Ch.						
Average							Distribute	BM	2:10	2:20	AP	3:00	3:08
Mental Age							Part Scores						
I. Q.							Total						
Educ. Age							Q-Score						
E. Q.							L-Score						
A. Q.							C. A.						
P. S. Grade							M. A.						
Av. Grade							I. Q.						
Acad. Apt.							Acad. Apt.						
Decile													
Total %ile	LO	1:23	1:30	MB	1:41	1:46							
Part %ile							Type	EM	4:00	4:12	MS	4:30	4:36
Local %ile											RO		
Item Anal.													
Over-print.													

FIGURE 1. Classification Slip for Cooperative Elementary Algebra Test, Form Z, Brownville High School, Used by the Educational Records Bureau in Processing Tests.

INTRODUCTION TO TESTING

swer rarely exists in the completely objective test, but in certain elementary-school tests in which the pupil is required to write out single-word responses, or in spelling tests where the handwriting of the pupil sometimes causes difficulty, the advice of the supervisor is needed on occasion. As may be imagined, the Bureau scoring departments have an extensive list of common alternate choices or responses which are or are not allowable as the answer to some test questions.

If the scoring clerk is a new one, every test is rescored by an experienced worker until the supervisor is reasonably sure that the new scorer is accurate. Thereafter, approximately one in five tests is rescored completely. Blue pencil is used for all checking operations. The number of correct responses and also the number of wrong responses, if the score is to be corrected for guessing, are then counted and recounted for each test booklet. Raw scores are changed to converted scores if converted scores are provided for the test, and this operation is also checked. Transferring scores from the inside of the booklet to the cover page is done and checked. It is especially necessary to check the counting, converting, and transferring of scores and all arithmetical operations since it is in these steps that errors large enough to affect seriously a pupil's score are likely to occur.

As mentioned in the preceding chapter, the responses to tests which are to be scored by machine must be recorded on separate answer sheets with special pencils. When the answer sheets are sent by the classification department to the machine-scoring department, they are first scanned to determine whether or not they will score correctly in the machine. Stray marks and dots are erased, and marks that are too light are darkened. They are then put through the

scoring machine by an operator, who reads each score from a dial. As a rule, the rate for scoring varies, depending on the tests, from about 300 to approximately 600 papers an hour. After the scoring is completed, the other processes are the same as those used with hand-scored tests.

Now the scores are ready to be distributed. The purpose of making a distribution sheet is to show the number of scores at each level. It will not identify the score of a particular pupil but will indicate the scattering of results within the group. The distribution gives a graphic picture of the standing of the class as a whole. A distribution is made by putting possible scores along the left side of a sheet, arranged from high to low, and putting a mark on the line opposite the appropriate score for each pupil's test. The completed distribution sheet for Brownville's beginning-algebra class would look something like that shown in Figure 2.

The next process is the computation of medians for classes of five or more pupils and also of quartiles for classes of fifteen or more pupils. This procedure will be described in more detail in the following chapter. The purpose of these computations is to determine the mid-point of the distribution, the point above which is the top one-fourth of the class, and the point below which is the lowest one-fourth of the class. For the distribution of scores of the Brownville algebra class, one-fourth of the scores fall below 61.8, which is indicated as Q_1 . The mid-point, or median, is 68.8, and the third quartile, Q_3 , above which is the highest quarter of the class, is 74.1. In Figure 2, the Q_1 , median, and the Q_3 points are entered below the distribution. When these processes have been completed, the test booklets are arranged in alphabetical order.

The material is now ready to be typed and assembled for a report to the school. Typed

SCHOOL *BROWNVILLE HIGH*DATE *6-5-50*

GRADE

*9*SCALED
SCOREYR. OF
STUDY*1 YEAR*

90

MR. CHRISTENSEN

88

86

84

82

80

78

II

76

III

74

III

72

III

70

III

68

III

66

II

64

III

62

II

60

II

58

I

56

II

54

I

52

*III**1 YEAR*

50

48

46

44

I

42

40

38

36

34

32

30-31

Total

47

Q3

74.1

MD

68.8

Q1

61.8

Range

37-79

- - - - End-of-year public-school median '50

FIGURE 2. Distribution Sheet for Cooperative Elementary Algebra Test, Form Z, Brownville High School, Used by the Educational Records Bureau.

INTRODUCTION TO TESTING

copies of the distributions are made and checked. Class lists showing pupils' names and test scores are typed and checked. As a rule, all material is typed in triplicate, two copies for the school and one copy for the Bureau's files, kept for reference and research purposes.

Certain practices in regard to class lists have proved helpful for the schools using central scoring. For instance, in reporting results for the fall program, the records for academic aptitude and reading tests for the same pupil are typed on one list. This makes it easy to study both types of data at the same time, and, when the records on achievement tests are available, the reading record will often shed a helpful light on the relation between academic aptitude test and achievement test results. In the preparation of reports for the spring program, secondary-school tests in English, literature, and spelling are customarily reported on the same list. Schools doing local scoring may wish to follow similar practices.

After checking and assembling are complete, medians and quartiles are indicated graphically on the distribution sheets and these sheets are clipped to the corresponding class lists. All material to go to the school is then assembled in a folder and an interpretative report is written. If the school wishes to have the test booklets returned for instructional purposes, they are sent back to the person in charge of testing.

A principal function which the Bureau performs is compilation of special norms for independent schools. Since the private school is considerably more selective in enrollment than is the public school, the usual published test norms are not entirely adequate for interpreting

the test performance of the independent-school pupil. The norms developed at the Bureau provide appropriate bases of comparison for the private school. Many public schools have found the Bureau norms based on independent-school populations quite useful, especially public schools having large proportions of college preparatory students. This particular function of the Bureau is mentioned at this point because of its relationship to test scoring. It is through its scoring service that the Bureau has been able to assemble a fund of information which provides probably the most extensive body of norm data anywhere in existence for selected students nearly all of whom are preparing for college.

To summarize: Although ease of scoring is implied in the term "objectivity," translating objective test performance into an accurate score demands careful attention to a number of details. Tests are frequently scored by the teachers who administer them. In order to insure accuracy in both understanding and applying scoring directions, special training is required. It is essential, too, that all processing procedures be checked in order to eliminate error. Scoring by a central agency affords a number of advantages, most important of which are economy in both time and money and confidence regarding accuracy of results. The scoring procedures followed by the Educational Records Bureau are described for the purpose of overall guidance in planning local scoring programs. It is emphasized that care and accuracy in scoring are as essential for valid test results as are care in selecting the instruments and professional competence in administering the tests.

HOW SHOULD TESTS BE SCORED?

SUGGESTIONS FOR FURTHER READING

1. Froehlich, Clifford P., and Benson, Arthur L., *Guidance Testing*, Chicago, Science Research Associates, 1948, pp. 49-50.
2. Lindquist, E. F. (ed.), *Educational Measurement*, Washington, American Council on Education, 1950, pp. 365-416.
3. Ross, C. C., *Measurement in Today's Schools*, New York, Prentice-Hall, Inc., 2nd ed., 1947, pp. 199-202.
4. Traxler, Arthur E., *Techniques of Guidance*, New York, Harper & Brothers, 1945, pp. 164-174.
5. Weitzman, Ellis, and McNamara, Walter J., *Constructing Classroom Examinations*, Chicago, Science Research Associates, 1949, pp. 70-80.
6. The separate manuals issued by test publishers contain detailed directions for the scoring of tests.



How Shall We Analyze and Interpret Test Results?

IF A TEST report were handed to a teacher not accustomed to using results of standardized tests, no doubt he would have difficulty in dealing with the data in a meaningful way. The array of figures, the use of terms having statistical connotation, the appearance of strange graphic symbols—all would add to his confusion. At once he would see that before he was able to determine answers to some of the questions he had concerning the performance of his class and the performances of individuals making up the class, he would have to become somewhat familiar with the terms and symbols used. This situation occurs frequently. After tests have been selected, administered, and scored, there remains the task of describing the results in terms which will permit comparison and analysis and which will be meaningful to all persons who will use the results either immediately or at a future time.

An attempt will be made in this chapter to describe in a simple, practical manner the terminology and techniques commonly employed in analyzing and interpreting test results. It is hoped that this unit may serve as a basis for group discussions in acquainting teachers with

the information needed to work intelligently with standardized test results.

TEST NORMS

A raw score in itself has little meaning. Usually, the first step toward interpreting test performance is to translate the numerical description of test performance represented by the raw score into terms that will indicate a comparison with or a placement among others who have taken the test. Probably the teacher will wish to compare each pupil with his own class and his own classmates. However, in addition, it is desirable to extend the size of the group with which comparisons are made. Any instructor working with small groups knows how difficult it is to judge the standing of his pupils accurately by any outside criterion. That is, while Sally may stand out well above the other pupils in her class in plane geometry, her teacher may still wonder whether Sally's work seems particularly good because the rest of the class is somewhat low or whether she is in fact exceptionally able in this field. If, after the administration of a standardized test, the teacher finds that the

HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

median of the class approximates the median for other groups of public-school pupils, he may find as well that Sally's percentile rating is very high, indicating that she would exceed a large proportion of most pupils taking the same subject. On the other hand, if the median attainment of the class as a whole is somewhat low, the top score may not be so far above the public-school median for the test as it is above the median for the single class in which the pupil is enrolled. It will readily be seen that test norms based on a wider group than the local class or school are equally useful when pupils transfer from one school to another. Since they give the standing of the individual in comparison with a large, stable group of pupils with generally similar training, they provide a useful adjunct to the more limited information provided by local comparison.

Test norms are developed through a process of standardization. This involves administering the test to a large, representative group of students at each age or grade level where the test will be employed. The results of such standardization may be reported in a number of ways. Among the more commonly used types of norms are the following:

1. PERCENTILE NORMS.

The percentile norm is perhaps the most widely used basis of interpreting test performance. The percentile describes the ranking or position of a particular score in terms of the percent of scores falling below the test performance in question. Thus, the pupil who achieves a percentile rating of 75 on a reading test displays a level of reading ability, as measured by the test, which surpasses that of 75 percent of the pupils included in the standardization or norm group.

Although the percentile is to be interpreted in terms of percentage placement, it is not to be confused with the percentage grade. A percentage grade designates what proportion of questions the pupil has answered correctly. On a twenty-question test, a pupil who answers eighteen questions correctly has a percentage grade of 90. If this pupil belongs to a class of 100 pupils and his was the third highest score, his percentile rating in his class is 97. If the lowest score in the class is eight of the twenty questions answered correctly, the pupil with that score has a percentage grade of 40 but a percentile rating of 1. Many tests constructed by teachers are planned so that a passing grade should be about 65 to 75 percent of the questions answered correctly. Usually, only a few pupils will be expected to have grades below this level. The average grade will be probably anywhere from 75 to 85. It may happen that the class average will be below the grade which the teacher has previously designated as the passing point. When percentile ratings are used, on the other hand, the average is always the fiftieth percentile and as many pupils are below that point as above.

Percentile ratings have some weak points. Chief among these is their inadequacy in differentiating accurately those pupils whose scores fall close to the center of the distribution. This is due simply to the fact that many more scores are found in this position than in either very high or very low percentile placements. The difference in *score points* between percentile ratings of 46 and 48 is actually much smaller than the difference in score points between percentile ratings of 2 and 4 or between ratings of 85 and 87, although the difference is 2 percent of the group in all three cases. However, in spite of this limitation, the percentile norm is prob-

INTRODUCTION TO TESTING

ably the simplest and most easily understood method of expressing the standing of an individual among other pupils at the same grade or age level with respect to achievement and ability.

2. GRADE NORMS.

Used frequently with achievement tests, particularly at the elementary school level, grade norms, or grade equivalents, are determined by testing large groups of pupils in each grade and computing average or median scores for each grade. Using the time of testing as a starting point, grade equivalents for scores falling between the averages are then assigned by interpolation or other statistical procedure. In such cases, the grade equivalent is often expressed as a decimal, assuming ten months for the school year. Thus, a grade rating of 2.7 means that the score in question is about average for pupils in the seventh month of the second grade. This type of norm assumes uniform progress throughout the grade range covered by the test. Actually, because of failure and retardation, there are not uniform differences in average chronological age from grade to grade, particularly in the intermediate and upper grades. In order to overcome this limitation of the traditional grade norm, another type, called modal-age grade norms, is used with some tests. These are based on the scores of pupils who fall within a limited age range at a particular grade. Thus, the effects of retardation or of acceleration are minimized and the modal-age grade equivalent provides a comparison of individual performance with a group making normal progress through the school, that is, where each increase of one year in grade level represents generally a year's increase in chronological age.

Grade norms seem to be popular with test

publishers because they are easy to explain. Since they are based on the familiar concept of grade classification, they assume meaning rather easily for the average teacher. They are particularly useful with achievement tests in the elementary grades, where grade classification is fairly well standardized throughout the country.

3. AGE NORMS.

The age norm is similar to the grade norm except that it is based on age level instead of grade level. Age norms or equivalents usually are based on average or median scores of representative groups of children at successive age levels independently of grade classification. In using age norms one assumes that academic maturity increases at a uniform rate with successive increase in chronological age. It is perhaps difficult to defend this assumption in view of varying opportunities to learn presented at different times in the calendar year. The age equivalent type of norm has its most common use in determining an intelligence quotient (I. Q.) or an educational quotient (E. Q.). In the case of I. Q., the ratio between mental age and chronological age is determined; if the E. Q. is sought, the ratio is found between educational or achievement age and chronological age. This description of educational and intelligence quotients is considerably oversimplified. Actually, a great deal could be written concerning the derivation, uses, and limitations of such indexes of brightness and achievement. The reader who is interested in this particular aspect of test norms is referred to the broader treatments provided in references suggested at the end of the chapter and in other books on measurement.

Another type of test norm, the Scaled Score, will be described later in the chapter. The three types which have been mentioned actually rep-

HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

resent the most frequently used bases of comparison.

It is necessary to observe certain cautions in using test norms. Some of these have been suggested in earlier chapters and are repeated here for emphasis:

1. A percentile rank, a grade equivalent, or an age equivalent assumes meaning only when something is known about the population upon which the norm is based. Use of published norms may be misleading if the students in a class, a school, or a regional group are notably different from the norm population with respect to general ability and educational opportunity.

2. Test norms interpret individual performance through comparison with others. While such comparisons are very useful, it is necessary sometimes to set absolute standards or goals by which to gauge progress and development. For example, pupils may be required to learn the spelling of all the words on a spelling list, or to master all the combinations in a certain multiplication table. In such instances, progress toward the expected goal of 100 percent mastery is usually measured by units of the expected goal rather than by comparison with the accomplishment of others. While it is of importance to know how the individual compares with others at various stages of growth and development, it is equally important to determine as nearly as possible his rate of development in terms of his own capacities and limitations. In this sense, each pupil presents his own "norm" or basis of comparison.

3. Test norms in one grade or age group are not directly comparable with those in another, and they may not be comparable from test to test. Hence, it is difficult to interpret accurately the amount and direction of growth and devel-

opment throughout the school year. Some test publishers have attempted to overcome this difficulty by preparing comparable forms of the same test standardized upon equally representative populations. However, this limitation has not been completely overcome.

To achieve adequate understanding of test norms and to analyze and interpret sets of test scores reported in terms of the various norms, it is necessary to know a little about elementary statistics. This is the tool by which a body of test data can be grouped and assembled so as to yield meaning. The next section is devoted to a brief discussion of the elementary statistical concepts used most frequently in analyzing and interpreting test scores.

WHAT ARE ALL THESE STATISTICS ABOUT?

Statistical methods originate in the desire to express a great deal of numerical information in the fewest and simplest possible terms. Statistics are never an end in themselves but are a means to clarification and simplification. Like the symbols of shorthand, statistical symbols appear mystifying to the uninitiated, but those needed in the use of test results are comparatively simple and easy to understand.

Suppose Mr. Burton, a teacher without formal training in statistics, has become a faculty member of a school which conducts regular testing programs and that his first contact with tests arises from a notice of this sort concerning a pupil who has been assigned to him as an advisee:

Pupil: Richard Jackson, Grade IX

Test: Algebra Achievement

Score: 69

His first reaction is "Fine," or perhaps "Not so

INTRODUCTION TO TESTING

good," and then he realizes that he is wondering, "69 what? 69 out of how many? Was it a difficult or an easy test? How well did Richard's classmates do?" His first move may be to look at the test and discover that there are 80 items and that, therefore, Richard has almost seven-eighths of them correct, if the score represents number of right answers, without the application of a correction for guessing.

It is still possible that the test was easy and that most of the class have scores as high as or higher than Richard's, so Mr. Burton goes to the ninth-grade class lists and looks up and down the list of 60 names until he sees that the highest score in the class was 78 and the lowest 39. Perhaps without being aware of the fact, he has ventured into the field of statistics and has obtained the *range of scores*. He has realized, too, in looking at the class list, that the score of 39 was a good deal lower than any of the others in the class and he begins to wonder where the class as a whole stands. The easiest way to find out is to get the average (which is exactly the same as the statistical term *mean*). Mr. Burton remembers from his elementary-school days that the average is the sum of the scores divided by the number of scores. So he adds up the scores, divides by 60 and obtains 62.1. Richard's score is considerably above average.

It now occurs to him, however, that the class as a whole may not have done as well as most ninth-grade classes on this test. So he decides to see whether the author of the test gives any information on what score the average pupil obtains. For this information he goes to the test manual and here he finds that the author says the national average for ninth-grade pupils is 63.0. Evidently the class average is very slightly below what has been determined to be the national "standard." However, Mr. Burton recalls

that the lowest score was considerably lower than any of the others and he wonders what effect that score has in lowering the class average. He decides to find out what score divides the upper half of the class from the lower half. In other words, what is the middle score in the class? He writes down on a sheet of paper all scores in the range from 78 to 39 from top to bottom, and then makes a little mark or tally for each pupil's score opposite that score on the page. When he is finished, his sheet looks something like the one shown here.

Now he starts at the bottom and counts until he comes to the 23rd tally, which is the last one opposite the score of 62. To complete the thirty cases which make up half the group of sixty, he needs an additional seven of the eight tallies at the score of 63. He then adds $\frac{7}{8}$ of a unit to the score of 63,¹ and the resulting point of $63\frac{7}{8}$ or 63.9 divides the scores of the sixty pupils exactly into two groups of thirty each.

78	/
77	
76	/
75	
74	//
73	/
72	/
71	
70	/
69	/
68	
67	///

¹ In this distribution the bottom 23 tallies extend through the score of 62 and up to the score of 63. In adding the needed $\frac{7}{8}$ of a unit to 63 we assume that the score of 63 extends from 63.0 to 63.9 and that the needed fractional unit should be added to the lower limit of the score (63.0). Many statisticians prefer to treat the limits of a score as extending from .5 below to .5 above the given score. Thus, the score of 63 would extend from 62.5 to 63.5, and, based on this assumption, the $\frac{7}{8}$ of a unit would be added to 62.5 instead of to 63.0.

HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

66 //
 65 ~~///~~ /
 64 ~~///~~ ~~///~~
 63 ~~///~~ ///
 62 //
 61 ///
 60 //
 59 ////
 58 //
 57 /
 56
 55
 54
 53 /
 52 /
 51 /
 50 /
 49 //
 48 /
 47 /
 46
 45
 44
 43
 42
 41
 40
 39 /

This middle point of 63.9 tells him that the class is very slightly above the national norm of 63.0 when the influence of the one low score is minimized. Mr. Burton has been dealing with statistics again! He has made a distribution of scores and has computed the *median*, or the point dividing the top half of the class from the bottom half.

Then perhaps he has another question. He wonders whether Richard happens to be in the top quarter of his class, and he counts down to Richard's score, 69, finding that he is eighth in

the class of sixty and so is clearly among the highest fourth, or the highest fifteen scores in the class. In doing this, he has seen that the scores of the class tend to cluster rather closely around the median, and it occurs to him that perhaps in describing the group as a whole he would need to express this tendency to cluster or to spread out. Without realizing it, he had begun to compute a measure of this spread by dividing the class into quarters, but stopped because he was concerned only with Richard's score, not with the scores of his classmates.

Let us now begin a somewhat more detailed and systematic discussion of the statistical concepts already mentioned and of a few others commonly used in reporting test results. First of all, we shall consider the distribution of test scores, that arrangement of scores in order from high to low which the teacher made as a preliminary step to finding the median for Richard's class. Owing in part to the way tests are constructed, many distributions of test scores tend to fall into a bell-shaped curve which more or less closely approximates a theoretical curve, generally called the *normal curve*. For practical purposes, all this statement says is that in distributions of test scores a larger proportion of the scores comes near the middle point and fewer and fewer at each side until they taper off at scores below the lowest attained in the group and above the highest recorded.

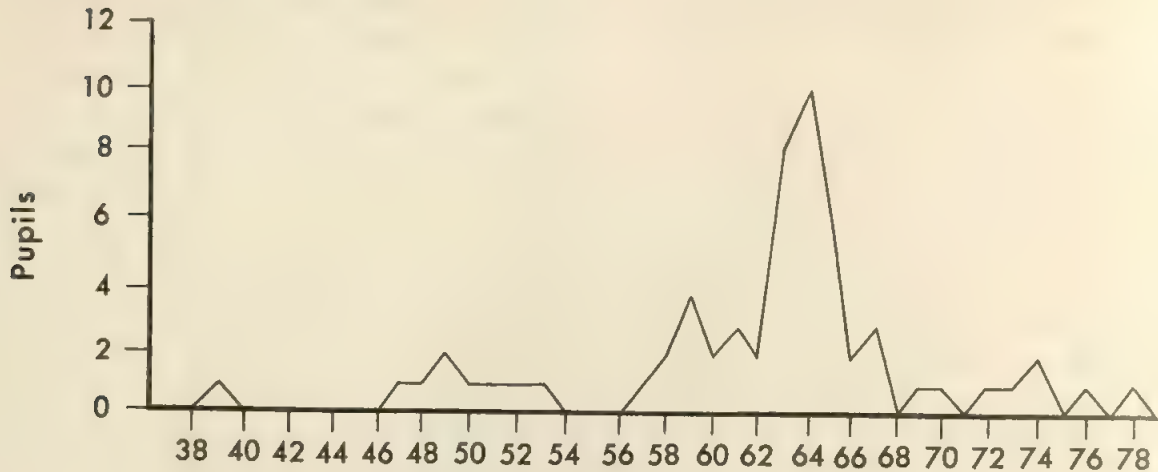
The distribution is likely to be rather irregular if there are only a few scores, but as more and more are added, the curve becomes more smooth and the "bell" figure more and more apparent. The distribution could be graphed with the scale of test scores running up and down, as in the teacher's tally sheet, but usually the test scores are placed at the bottom of the graph and the number of cases is put on the left side. Thus

INTRODUCTION TO TESTING

the distribution of test scores for the sixty cases is shown in the first diagram on this page.

percent of the scores fall. Actually, the "normal" curve is a theoretical concept, and it rarely oc-

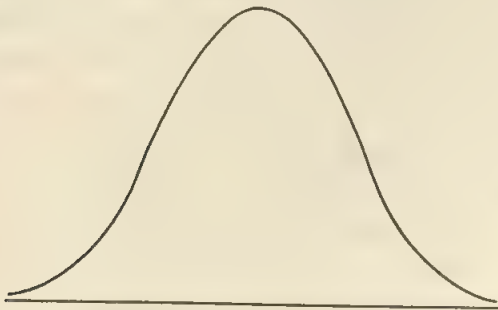
The distribution of test scores for the sixty cases looks like this:



SCORES

The normal frequency curve is a smooth curve, as shown in the second diagram.

The normal frequency curve is a smooth curve like this:



As one can see, Mr. Burton's distribution is not normal, but it would probably look somewhat like the smooth curve if enough cases were added. The normal curve is symmetrical as well as smooth. That is, it may be divided by a line through the middle so that one half is simply the mirror image of the other half. In a normal distribution, both the mean and the median are at this middle point above and below which 50

curs in graphing a distribution of scores. Ordinarily distributions are not exactly symmetrical, and the mean and median do not precisely coincide. If the test questions are too easy for the group tested, the scores will be grouped at the high end of the scale; if the items are too difficult, the scores will tend to accumulate at the lower end, rather than near the center. Most school classes will be like the group of sixty studied by Richard's teacher, with the majority of the scores grouped about the median and with considerable irregularity shown as one enters on the graph the scores which are at a distance from the median.

The *median* is by definition the point above and below which half the scores fall. It will be stable as a measure of central tendency even if the exact placement of the highest and lowest scores changes considerably. The *mean*, or average, on the other hand, is influenced by the size of each score, and so will be considerably in-

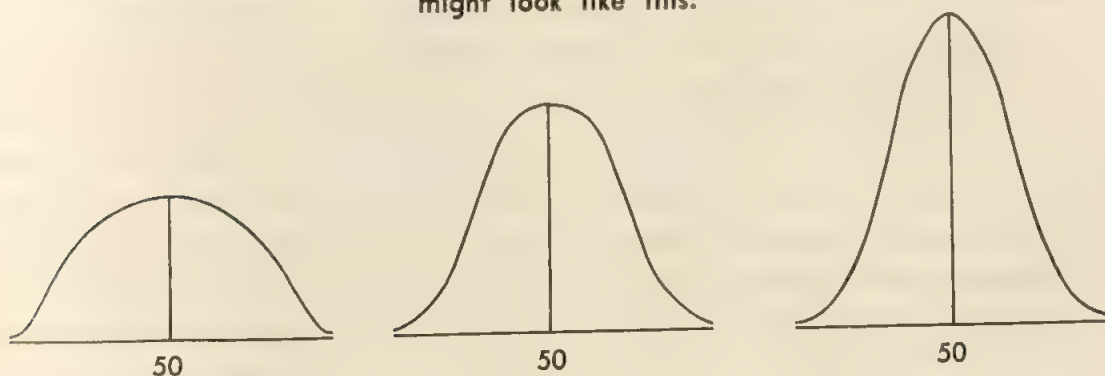
HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

fluenced by the extreme cases. Choice between the median and mean as descriptions of the central tendency of the group of scores is made on the basis of the use to which the measure will be put. While the median score is easy to secure and is used commonly in describing the performance of a group, the mean takes into account the extremes of achievement as well and is often needed as a basis for more extensive statistical analysis of the scores.

Not all symmetrical curves with the same

ever, so subject to chance are the exact positions of these extreme cases that they cannot always be used to describe accurately a whole class. We therefore look for a more reliable measure of spread. Let us consider the measure which Mr. Burton started to compute but did not complete. This measure is one which includes the middle 50 percent of test scores and is called the *interquartile range*. It is obtained by counting down from the top to find the point above which is the top quarter of the cases and count-

Three "normal" curves having the same mean and different variabilities might look like this:



central tendencies have the same shape. Consider, for example, the three shown on this page. The mean or average of each curve is at 50, yet there is a great deal of difference in the extent to which the scores represented by the three curves are spread out around the mean.

We need a measure in addition to the mean or median to describe these curves. We want to describe or summarize in one term their characteristic "spread-outness," or "pushed-togetherness," just as we characterized in one measure their central tendency (the median or mean). The first possibility that occurs to us is to use that first statistical measure which Richard's teacher obtained, the range of scores, represented by the highest and lowest scores. How-

ing up from the bottom to find the point below which is the lowest quarter. The point which separates the highest fourth from the rest of the scores is called the *third quartile* or Q_3 , and the corresponding point marking off the bottom fourth is called the *first quartile* or Q_1 . The difference between these two, $Q_3 - Q_1$, includes the middle 50 percent and is known as the *interquartile range*.

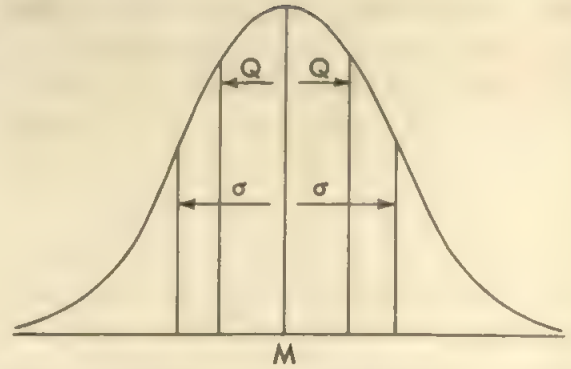
The computations required are similar to the computation used by Richard's teacher in securing the median for his distribution. There are 60 pupils in this class, and one-fourth of this number is 15. Starting at the top, we count down through the score of 66 where we have 13 cases and need two of the six opposite the score of 65.

INTRODUCTION TO TESTING

We therefore subtract $2/6$ of a point from 66.0 and have Q_3 equal to 65.7. The twelve lowest cases bring us up through the score of 58 and we need three of the four opposite 59, so we add $3/4$ of a unit to 59.0 and obtain 59.8 as Q_1 . The distance from Q_3 to Q_1 then is equal to the distance from 65.7 to 59.8, or 5.9 score units, and 50 percent of the class falls in this interval extending over 5.9 score points. Usually this interquartile range, expressed as $Q_3 - Q_1$, is divided by two to obtain the statistical measure designated by the single term Q (without the subscript). In the "normal" distribution, the distance between the median plus Q and the median minus Q also contains the middle 50 percent of the cases. This distance coincides with the interquartile range. We see, therefore, that in describing a distribution we use both a *point* (the median) and a *distance* (Q) which is measured in both directions from the median.

There are other measures of the amount of "spread-outness" or scatter (in formal statistical terminology this characteristic of a distribution is called *variability*). The measure used most frequently in statistical work is the *standard deviation*. This measure is designated by the small Greek letter σ and is sometimes called by the name of that letter, sigma. The standard deviation is a distance (as Q is also), but it is always measured from the mean, never from the median. Furthermore, it is a greater distance than Q , for the area under the normal curve between the mean plus 1σ and the mean minus 1σ includes 68.26 percent, or approximately two-thirds, of the cases in the entire distribution. The relation between these two measures of variability may be illustrated by reference to a normal curve on which both are shown.

It is outside the purpose of this book to de-



scribe in detail the procedure necessary to obtain the standard deviation.¹ This measure is the square root of the mean of the squares of the distances of each score from the mean of the distribution.

One reason for using the rather complex standard deviation when the much simpler Q is available is that the standard deviation is necessary for further statistical analysis. For example, in finding correlations, the computation of means and standard deviations is usually a necessary step. With small groups, however, the standard deviation is subject to the same criticism as the mean, namely, that it is greatly affected by extreme scores. Since the median instead of the mean is used ordinarily as a measure of central tendency with groups of moderate size, such as a single class, it is convenient to use Q instead of the standard deviation as a measure of variability.

Measures of central tendency and measures of variability are employed not only in analyzing and interpreting local results but also in describing standardization populations upon which published test norms are based. It would be preferable, of course, if the information about the central tendency and the variability of the norm group could in some way be incor-

¹ The reader is referred to any one of the suggested readings at the end of this chapter.

HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

porated into the test score itself. Various test makers have tried in several ways to work out such a score. One of the most successful of such systems of scores is the plan of Scaled Scores used with the Cooperative Achievement Tests. The makers of these tests decided, "We are going to define our standard group for a given subject as high-school pupils who are completing their study of the subject and who have had the usual kind and amount of instruction. These pupils shall be so selected that the average I. Q. of the group will be 100. Now we are going to take all of the distributions of test scores from the standard group and change these scores into a common scale in such a way that all the mean scores are 50. We will multiply or divide the standard deviations by quantities such that they all become equal to 10. Then we will be able to express each test score in terms of its distance in tenths of standard deviation units from the mean."

From the resulting system of Scaled Scores each single score has incorporated in it information about its distance from the mean of the standardization group in terms of the variability of that group. Thus, a Scaled Score of 62 is 1.2 standard deviations above the mean of this defined standard group, and a Scaled Score of 44 is 0.6 standard deviations below the mean. The system also results in equivalent scores from test to test and from one part to another of the same test. That is to say, a Scaled Score of 57 in algebra is 0.7 standard deviation units above the mean and is equal to, or has the same meaning as, a Scaled Score of 57 in physics or American history or any other subject.

The course of study itself for which the test is intended must be carefully described and defined. It is particularly important to state definitely what shall be considered the end of each

course of study, since the point of reference (Scaled Score of 50) is defined in terms of the average score at the *end* of the course. For English, literature, and other subjects continuing throughout the high-school years the point of reference is at the end of four years of study. For languages the reference point is at the end of two years of study, and for other subjects, such as plane geometry and ancient history, at the end of one year of study in the high school. The authors have also stated in what grades these courses were taken by the standard group. For example, a Scaled Score of 50 in any foreign language is the average score for pupils at the end of two years of study in Grades 10 and 11.

As indicated previously, the Scaled Score system is based on the test performance of public-school pupils whose average I. Q. is 100. In the actual school situation, however, the average I. Q. of public-high-school pupils is likely to be higher than 100 (generally about 103 to 105) since the very dull pupils drop out or are placed in special schools, while the average and brighter pupils remain. In most public high schools, therefore, the average Scaled Score for a given test will be somewhat higher than 50.

It has already been pointed out that one of the limitations in use of percentile norms is the fact that the difference in score points between percentile ratings near the center of the distribution is actually much smaller than the difference between percentile ratings at the extremes. Scores expressed in terms of standard deviation units, such as Scaled Scores, are not open to this criticism. It is possible and often advisable to equate percentiles to Scaled Scores or another standard deviation unit, by referring to the normal curve for which theoretical frequencies are known.

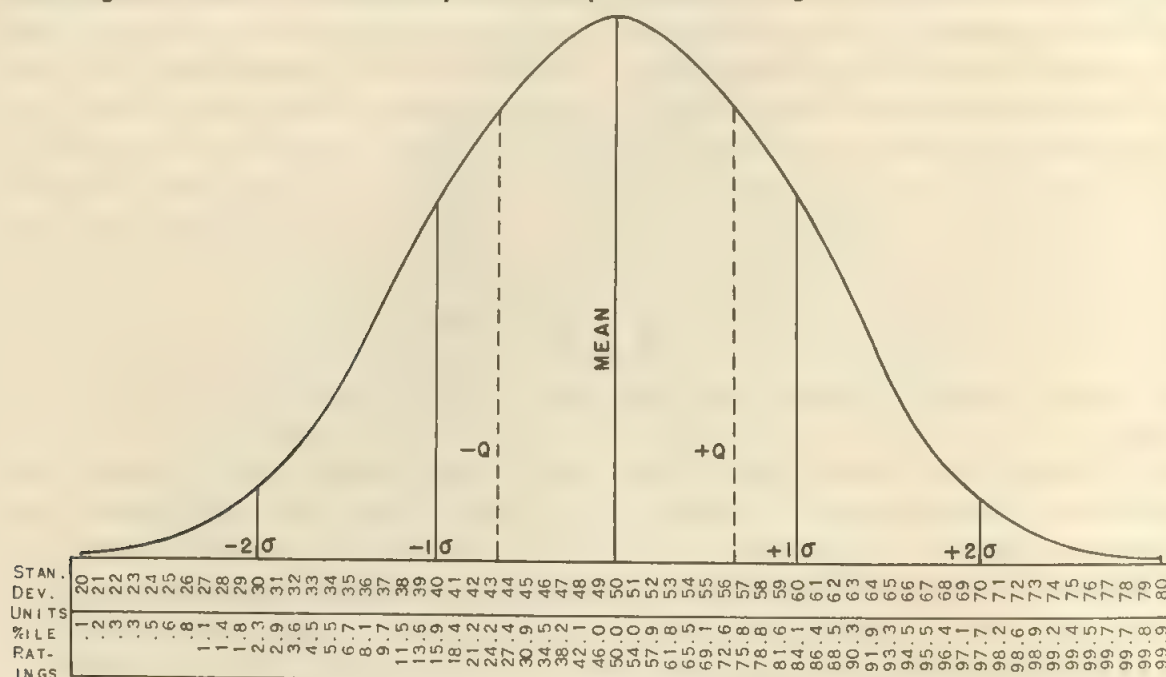
The accompanying diagram shows a normal

INTRODUCTION TO TESTING

curve with the base line marked off in units of both Scaled Score and percentile ratings. The relationship of each to the mean and to standard deviation units is indicated by perpendiculars to the base line, appropriately labeled. We may see from the diagram that while 4 percent of the scores fall in the 0.1 standard deviation unit (or the one Scaled Score unit) between 49

class list, one may feel that in some way he should be able to say to what extent the same pupils have high scores or low scores on both tests—that is, whether pupils scoring high on the one test tend to score high also on the other. For example, Susan has the highest English score but is only sixth highest in reading comprehension, while Jim has the highest reading

This diagram shows the relationship between percentile ratings and standard deviation units:



and 50, only one-half of one percent of the scores fall in the 0.1 unit between 29 and 30.

CORRELATION

There remains at least one more statistical term about which an intelligent user of test results should know a little, and that is *correlation*. Suppose, for a given class, scores are available on two tests, such as reading comprehension and English usage. The class medians and the interquartile ranges for the two tests have been determined. Looking up and down the

score and is fourth in English usage. If the class has thirty members, of course, these two are near the top in both tests. Such comparisons for each of thirty pupils become tedious, and there should be some method of telling in simple terms to what extent the scores on two tests vary together—that is, to what degree high scores on one test are accompanied by high scores on the other test. The statistical term for the measure expressing this degree to which pupils are ranked in the same way by the two tests is the *coefficient of correlation*.

Coefficients may vary from +1.00 to -1.00.

HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

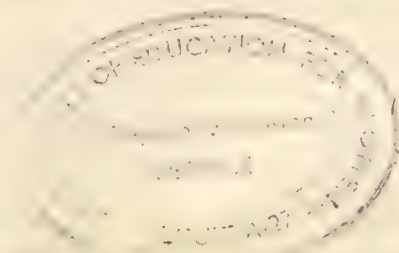
With a correlation of $+1.00$, the person with the highest score on one test also ranks first on the other test, the person who is second in one test also is second in the other, and so on, with the pupil who is lowest in one test also lowest in the other. Where the correlation between the two tests is -1.00 , the person who ranks first on one test is last on the other test, and so on. If there is no relationship between the series of test scores, that is, if knowledge of a pupil's rank on one test tells nothing about his rank on the second test, the correlation is 0.00 . Actually, it is unusual to find correlations of $+1.00$ or -1.00 , and when ability and achievement scores are being dealt with, negative correlations are seldom found. In other words, there is some tendency for high achievement in one field to be accompanied by high achievement in other fields. Negative correlations might be found, however, between such variables as body weight and speed of running, and a zero correlation might be found between finger length and school marks in history.

In this connection, it should be stressed that the correlation coefficient tells us nothing about the *nature* of the relationship between the two variables. There may be a cause-and-effect relationship, or variables may be related in some way to a third variable. Furthermore, the size of the correlation does *not* express the percentage of relationship. The interpretation of correlation coefficients depends largely on the material under consideration. A correlation of $.60$ between school marks and intelligence test scores, for instance, is considered to be rather high, but such a correlation between two forms of the same test would be low. A correlation of $.30$ between marks and intelligence would be low, while such a correlation between pupils' heights and I.Q.'s would be astonishing. Inter-

pretation of correlation coefficients must, therefore, depend heavily on knowledge of the amount of relationship usually found to exist between variables of the sort in which we are interested.

The full value of correlation coefficients would not be realized if they were used only to show relationships existing between test scores which were obtained on a single class. Suppose, for example, that a certain arithmetic test is given to a class beginning algebra and then after a year of instruction the same class is given a test in elementary algebra. The test scores yield a correlation of $.80$. This fact is of academic interest only, unless we see a way of making use of it. For instance, this correlation suggests that by giving the arithmetic test at the beginning of the course we can predict generally that most of those who score low will also score low in the final algebra test, while most of those who do well on the arithmetic test will later do well on the algebra test. Therefore, we would appear to be justified in assigning pupils to fast or slow sections in algebra on the basis of scores on the initial arithmetic test, or perhaps in requiring a general mathematics course as a prerequisite to algebra for those scoring low on the first test.

The use of test results to predict some future achievement depends on knowing how these test results correlate with a measure of that achievement. However, one must be exceedingly cautious about generalizing from an obtained correlation. The size of a correlation depends in part on the group tested and so should be applied only to similar groups. A correlation obtained between two measures of a large group of ninth-grade public-school pupils does not justify the assumption that the same relationship will exist between these same two measures for



INTRODUCTION TO TESTING

airplane mechanics or for medical students or even for ninth-grade independent-school pupils. It may be justifiable to assume that a rather similar correlation will be obtained from testing another large group of ninth-grade public-school pupils from high schools similar to those attended by pupils of the first group. Even if the measurements are repeated on the original group, exactly the same correlation as that obtained first would probably not be found. That is why a statement of the *probable error* of a correlation is almost universally reported with every coefficient.

The probable error (P.E.) shows the range within which correlation coefficients would be expected to fall if many similar groups were tested and correlations computed. Thus, if a coefficient with its probable error is written as $+.55 \pm .02$, and numerous correlations were then computed on similar groups of the same size, one would expect in 50 percent of the correlations to get coefficients between $+.53$ and $+.57$. One-fourth of the correlations would be expected to fall above $+.57$ and one-fourth below $+.53$. Almost all correlations would fall between the obtained correlation and four times the probable error on either side of it, that is, between $+.47$ and $+.63$. One of the important functions of the probable error is to tell whether a correlation coefficient is significantly greater than zero and, therefore, indicates a true relationship or whether there may be no real relationship but by chance a coefficient other than zero was obtained. A correlation may be said to be significantly greater than zero if it is four or more times as large as its probable error.

The special kinds of correlation called reliability coefficients and validity coefficients have already been mentioned in an earlier chapter. It will be recalled that these refer respectively

to the consistency of measurement yielded by a test and to the trueness with which it measures what the test is purported to measure.

Such a summary of statistics as we have presented here is of necessity rather superficial, but perhaps it will prove helpful especially to those who are confronted for the first time with statistical terms. There is nothing essentially mysterious or difficult about statistical concepts; even the most elaborate methods are designed for the purpose of summarizing and simplifying the information contained in many test scores. An acquaintance with elementary statistical methods will enable teachers to use quantitative information about their pupils more intelligently and more effectively. Statistical methods should be put in their proper place with other methods directed to the same end, namely, to serve as tools for the teacher with the guidance point of view—a means to a better individualized education.

ANALYSIS OF TEST RECORDS

When tests are scored by a central agency such as the Educational Records Bureau, results are usually reported to schools in such form as to facilitate interpretation and use by the school faculty. If the tests are scored locally, there is the task of deciding about the method or procedure for gathering the test data so that each teacher may make certain analyses for her class.

It may be helpful in some situations simply to use the test booklets or the answer sheets upon which are entered the individual scores and percentiles. This procedure has some advantage in that wrong responses are identified easily as the results are explained to the pupils,

HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

and there may be instructional value in pointing out correct responses. However, the teacher will not be able to get an accurate picture of group performance from study of the individual answer sheets, and it is true that going back over the test questions will render useless for some time the particular form of a particular test which may be employed.

Analysis of group performance is an important aspect of interpreting results, not only because of the clues it may yield regarding strengths and weaknesses in the program of instruction, but also because of the assistance provided in interpreting individual performance. It is important, for example, to know how far above average or how far below average a particular pupil may be *with respect to his own group*.

The distribution of scores described in the first section of this chapter is one means of studying group performance. Consider, for example, the score distribution shown in Figure 3. The scores of 214 Grade 10 pupils on the Cooperative English Test C₁: Reading Comprehension, Form T, form the basis of this distribution. The median score in the distribution, the Q₁ and the Q₃ distribution points, and the total range in Scaled Score from highest to lowest have been computed and entered below the distribution. The median is shown graphically by the short horizontal line near the center of the distribution, and the range of scores between the Q₁ and Q₃ points is marked off by a vertical line adjacent to the distribution.

Here, then, is pictured for the teacher an indication of the range of reading comprehension skills among the Grade 10 pupils in this school. Also, it is possible to perceive more accurately the variation from the mean or average represented by high and low scores.

Cooperative English Test C₁ Reading Comprehension, Form T

TOTAL SCORE	
Summitville High	Grade X
Scaled Score	Frequency
90-91	
88-89	1
86-87	
84-85	
82-83	
80-81	1
78-79	3
76-77	3
74-75	2
72-73	1
70-71	1
68-69	3
66-67	4
64-65	6
62-63	8
60-61	14
58-59	15
56-57	17
54-55	20
52-53	9 —Md.
50-51	9
48-49	14
46-47	16
44-45	9
42-43	10
40-41	12
38-39	9
36-37	6
34-35	5
32-33	2
30-31	5
28-29	2
26-27	2
24-25	2
22-23	2
20-21	1
18-19	
Total	214
Q ₃	59.1
Md	52.2
Q ₁	43.1
Range	21-88

FIGURE 3. Scaled Scores and Frequencies for Summitville High School Grade 10 Cooperative English Test C₁: Reading Comprehension, Form T.

INTRODUCTION TO TESTING

Other information about group performance can be obtained, using the score distribution as the basis of analysis. For example, the tenth-grade pupils in the Summitville public school can be divided into two groups according to course objective. Such division finds 85 of the total of 214 classified as business and general students, while 129 are in the college preparatory curriculum. When the scores for these two groups are distributed separately, as shown in Figure 4, notable differences can be observed. As one would suppose, the median score for the business and general group is considerably lower than the college preparatory median. However, there is overlapping of distributions. A few of the students in the business and general group have scores which surpass the college preparatory median. On the other hand, several college preparatory students are below the median of the business and general group. In fact, three-fourths of the business and general students appear to be more skilled in reading comprehension than are the two lowest students in the college preparatory group.

Further analysis of the results for the two groups can be made by plotting into the distributions the "national" median supplied by the test publisher for this particular grade level. This is shown in Figure 4 by a broken line drawn through both of the distributions. It will be seen that the median for the business and general group is somewhat below the publisher's norm, whereas nearly three-fourths of the college preparatory pupils have scores which surpass the "national" median.

This brief discussion suggests some of the uses which may be made of the distribution of scores in analyzing test performance. Some publishers provide forms for preparing the distribution as part of the supplementary materials

Cooperative English Test C ₁ Reading Comprehension, Form T			
TOTAL SCORE			
Summitville High	Grade X		
Scaled Score	Business and General	College Preparatory	Total
90-91			
88-89		1	1
86-87			
84-85			
82-83			
80-81		1	1
78-79		3	3
76-77		3	3
74-75		2	2
72-73		1	1
70-71		1	1
68-69		3	3
66-67		4	4
64-65		6	6
62-63		8	8
60-61	1	13	14
58-59	2	13	15
56-57	4	13	17
54-55	4	16	20
52-53	1	8	9
50-51	3	6	9
48-49	11	3	14
46-47	6	10	16
44-45	3	6	9
42-43	9	1	10
40-41	9	3	12
38-39	8	1	9
36-37	5	1	6
34-35	5		5
32-33	1	1	2
30-31	4	1	5
28-29	2		2
26-27	2		2
24-25	2		2
22-23	2		2
20-21	1		1
18-19			
Total	85	129	214
Q3	48.9	62.2	59.1
Md	42.3	57.2	52.2
Q1	36.9	51.8	43.1
Range	21-60	30-88	21-88

FIGURE 4. Scaled Scores and Frequencies for Business-General Course and College Preparatory.

HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

included in packages of tests. Figure 5 shows one such form. This particular illustration makes up one section of the class record form provided by the World Book Company² as part of the materials for the Iowa Silent Reading Test.

It is possible to picture more accurately the individual placement of each pupil in a distribu-

DIRECTIONS FOR USING THE DISTRIBUTION TABLE

To make a distribution of median standard scores, first enter the proper class intervals in Column 1. A class interval of five points is recommended.

For each measure to be distributed, make a tally mark opposite the interval in which it falls. (Thus, for a score of 53, for example, a tally should be entered opposite the interval which includes this value, probably 50-54.)

When all the measures have been tallied in this fashion (|||| |), count the tallies opposite each interval and write the number (frequency) in the third column.

The median,¹ or middle, measure may be determined in several ways. One method is to arrange the measures or scores from highest to lowest, and count to the middle measure, if there is an odd number of cases; if there is an even number of cases, the average of the two middle measures is the median. A second and statistically more accurate method of finding the median is as follows:

1. Divide the number of scores by 2.
2. Add the frequencies from the lowest score up to but *not including* the interval that contains the middle score.
3. Subtract this sum from half the number of scores.
4. Multiply this difference by the number of points in the score or class interval of the distribution.
5. Divide this product by the number of pupils in the interval containing the middle case.
6. Add this quotient to the lower limit of that interval²—that is, of the interval containing the middle case.
7. This sum is the median.

¹ For more detailed information concerning the computation of medians and the analysis of test scores, see *Test Method Help No. 4, "Statistical Methods Applied to Test Scores,"* published by World Book Company.

² While practice differs somewhat, it is preferable to consider the lower limit of the class or score interval as being five tenths of a point below the value printed in the table (e.g., 74.5 rather than 75).

MEDIAN STANDARD SCORE DISTRIBUTION		
CLASS INTERVAL	DISTRIBUTING MARKS	FREQUENCY
	Total	
	Median	

FIGURE 5. Publisher's Distribution Sheet for Iowa Silent Reading Test. (Used with permission of World Book Company.)

tion of scores by entering the name of the pupil in the interval where his score falls rather than entering just a tally. A procedure similar to this is recommended, for example, by the publisher of the Stanford Achievement Tests. Rather than entering the names of the pupils, a code number is assigned to each and this number is entered on a chart in the interval opposite the score made by the pupil, an entry being made for each part score of the achievement battery. The

² See Appendix.

class analysis chart supplied by the publisher with a sample class record entered is shown in Figure 6.

It will be seen that this chart presents a sort of profile of the median scores obtained by the class. Thus, it is possible to observe readily the relatively strong and weak achievement areas

found in the group. The class appears to be best prepared in the knowledge and skills measured by the elementary science and arithmetic reasoning tests. The lowest median achievement occurs in word meaning. Other than the fairly large differences between the word meaning and the science and arithmetic reasoning median, the profile shows rather consistent performance on the parts of the achievement battery, the mid-scores falling generally between grade equivalents of 4.0 and 4.3. The widest range of

TO USERS OF SEPARATE ANSWER SHEET EDITION. Note that the arrangement and the numbering of the subtests in the Separate Answer Sheet edition differ from the arrangement and the numbering followed in the Class Analysis Chart. Be careful to enter the scores for each test in the column headed by the name of that test.

CLASS ANALYSIS

1	2	3	4	5	6	7	8	9	10	11
AGE EQUIV.	GRADE EQUIV.	EQ. SCORE	TOTAL (AVE.)	1. PAR. MEAN.	2. WORD MEAN.	AVE. READ.	3. LANG. USAGE	4. ARITH. REAS.	5. ARITH. COMP.	AVE. ARITH.
		79 or above								
16-0	11.0	78								
15-10	10.9	77								
15-7	10.6	76								
15-4	10.3	75								
15-0	10.0	74					17			
14-9	9.8	73								
14-6	9.5	72					16			
14-3	9.3	71								
14-0	9.0	70								
13-9	8.8	69								
13-6	8.5	68								
13-3	8.3	67								
13-1	8.1	66								
12-10	7.8	65								
12-8	7.6	64								
12-5	7.4	63			17					
12-2	7.2	62								
12-0	7.0	61								
11-10	6.8	60								
11-7	6.6	59								
11-5	6.4	58	17			17	14			
11-3	6.2	57					15, 20			
11-0	6.0	56			16		10	23		
10-11	5.9	55								
10-10	5.8	54								23
10-8	5.6	53								
10-6	5.5	52								
10-5	5.4	51	16	7, 17, 23	10	16, 23				
10-3	5.2	50			20, 23		23	12	23	12
10-2	5.1	49								
10-0	5.0	48	20, 23			10	5			
9-11	4.9	47	10	13				13, 15, 17, 20	8, 12, 20	20
9-9	4.8	46		5, 15, 16		5, 7, 20				15
9-8	4.7	45	14	10, 12	5			1, 16, 22	15	8
9-7	4.6	44				13				1
9-6	4.5	43	1, 7, 12, 13				1		1	
9-5	4.4	42	5, 15	14, 20	1, 9	1	12	5, 8, 14	10	5, 17
9-3	4.3	41		1, 22	13	15	19		18	13, 14, 22
9-2	4.2	40			7	12	13, 22		5, 9, 14	10, 16
9-1	4.1	39	8			14	8, 9			
9-0	4.0	38		8	21	9		10	3	9
8-11	3.9	37							17, 19, 22	3, 18
8-9	3.8	36	22			22		3, 7, 9, 11, 21		19
8-8	3.7	35	19		3, 14, 15	8, 21	21			7, 11
8-7	3.6	34	3, 9	9, 11	4, 12			2, 14, 18	4, 13, 16	4, 21
8-6	3.5	33	4, 11, 18, 21	18	6	3, 4, 11	3	18	7, 11	2
8-5	3.4	32			11					
8-4	3.4	31			8, 19		18		21	
8-3	3.3	30	2, 6	2, 3, 4, 21		6, 18, 19	2, 11	6		6
8-3	3.2	29		19		2			2	
8-2	3.1	28		6	2, 10, 22		6			
8-1	3.0	27							6	
8-0	3.0	26								
7-11	2.9	25								
7-10	2.9	24								
7-9	2.8	23					4			
7-9	2.7	22								
7-8	2.6	21								
7-7	2.6	20								
Median										

FIGURE 6. Class Analysis Chart for the Stanford Achievement

CHART

Stand. Ach. Test: Inter. and Adv. Compl.

12	13	14	15	16	17	18	19	20
6. LITERATURE	7. SOC. STUD. I	8. SOC. STUD. II	AVE. SOC. STUD.	9. ELEM. SCI.	10. SPELLING	Eq. SCORE	CHRON. AGE	IQ
						79 or above		129 or above
						78		128
						77		127
						76		126
						75		125
						74		124
						73		123
						72		122
						71		121
						70		120
					17	69		119
						68		118
						67		117
						66		116
						65		115
						64		114
						63		113
						62		112
				17		61		111
						60		110
17, 20				10		59		109
						58		108
						57		107
	7, 17	16				56		106
						55		105
						54		104
1, 7			16			53		103
			17	23		52		102
			7			51		101
	12, 16			1, 13	20, 23	50		100
13, 14		17			14	49		99
			12			48		98
		10			10	47		97
16		7, 12		7, 16		46		96
				8		45		95
	13	14	10	5, 14, 20		44		94
			13			43		93
5, 15		3, 13	14	12		42		92
10, 23	1	23			4, 5, 8	41		91
	6, 10, 14, 20	15	20			40		90
		22, 21	23		12, 16	39		89
0, 12	19	5	1	15	3, 13, 15	38		88
	23	8	6, 15, 21			37		87
21		18	3, 19		22	36		86
12	4, 22		5		1, 18	35		85
11	11, 15, 21	1, 2	4, 8, 11, 22	19		34		84
		4, 6, 9, 11, 19, 22	2, 18		19	33		83
3, 22	2, 5		9	22	11	32		82
				9		31		81
						30		80
	3, 8, 9, 18			4	7	29		79
					2	28		78
4, 9, 18				3, 11	6, 9	27		77
2, 6				2, 6, 18, 21		26		76
						25		75
					21	24		74
						23		73
						22		72
						21		71
						20		70 or below
							Median	

concentrate on the language map and spelling tests. In language maps, since there are no separate tests for people who have the linguistic highest and lowest grade levels, spelling achievement is the best single test. A grade equivalent of 1.4 up to 4 placement is best.

In addition to providing information about group performance, a chart such as the grade level chart of individual performance. For example, it is useful to identify in pencil on the chart pictured in Figure 5 the individuals who are contributing to the group. The graph indicates the number of students having at the time in total average and to show if the low ability is well. In fact, the graph might be converted to placement in a higher grade, since it is the evidence of lower achievement in relation to the individual comparison. Each number 1 on the chart means a consistently low on the low level path of the test, that is, a low number from the low's achievement experience. The evidence for his class. On the first column of this chart he goes to show the teacher.

The three columns at the left side of the chart, age, linguistic, grade equivalent, and reading score, are in effect a table of achievement at a given level, age and grade equivalent, and the best expected score. First, if a positive is shown at one time the placement of an individual is not well with respect to the class and the corresponding age and grade scores.

Clearly, then, the development of group charts in checking low scores from the analysis of both group and individual analysis.

There are other ways in which the individual test results for a class may be analyzed. For example, chart recording strengths and weaknesses of instruction may be shown from studying the responses to separate items or questions

in the test, might be looking for evidence of right and the evidence of wrong responses to each item, suggest, for example, that reading lists are that of the class in a particular test taken by the pupils yielded the following results:

Item	Right		Wrong	
	No.	%	No.	%
10	10	100	11	110
11	10	100	1	10
12	10	100	10	100
13	10	100	11	110
14	10	100	11	110
15	1	10	10	100

The table suggest that item 10 is correct in relation to group. Item 11 was found by the group itself, whereas the achievement failed to be question 10 had been marked by about 11 of the group. Item 12 is correct, but possible achievement, the learning the test level. Another comparison of the table might be that item 13 is much too easy to be included in the test, while item 15 may be too difficult, may be as ambiguous item as item 14 is found wrong.

Comparison of percentages, comparison of individual responses, comparison of individual responses will be used for the first time. It is clear that achievement, however, he would that language is not by itself, might be learned from the first time. The student analysis which have been suggested can be analyzed by most teachers and will give an insight to study and interpretation of results.

The teacher will wish to have some complete record of the test scores for the class. Such a class list, showing the names of all pupils and the scores and percentage marks obtained by each, provides a convenient work sheet for making various chart analyses and can serve as a master for record against which to check individual individual. (See: Supplemental)

INTRODUCTION TO TESTING

materials accompanying most tests will include a class record form, typical of which is the one illustrated in Figure 7. This record has been prepared for use with the Iowa Silent Reading Test. Spaces for additional names are found on the reverse side of the form, which is not shown in the illustration. It will be noted that one column in this form is provided for M.A. or I.Q. This practice is in keeping with the trend toward consideration of aptitude level as achievement results are studied.

The class list is a useful device for bringing together several test scores, so as to facilitate comparisons between results of separate tests and to insure that results on any one test are not regarded as single items of information. In order to illustrate the use and interpretation of the class list data in this respect, reference is made to a typical report of test results drawn from the files of the Educational Records Bureau.

The class list illustrated in Figure 8 gives the records for an eleventh-grade class at the Larchmont School on the American Council Psychological Examination and the Survey Section of the Diagnostic Reading Tests. These were administered as part of a fall testing program. Part and total scores are shown for the academic aptitude and reading tests. Additional information is furnished by the inclusion of the Otis equivalent mental ages and intelligence quotients which correspond to the total scores on the American Council test. These results are derived from equating the Self-Administering Test of Mental Ability with the successive college-freshman editions of the American Council test. The Otis equivalent I.Q.'s obtained in this way are an approximation of the intelligence quotient, useful for schools accustomed to dealing with this type of record of mental ability.

The quantitative (Q) and the linguistic (L) scores yielded by the American Council test give a somewhat diagnostic picture of the ability of the individual and may be helpful for predicting success in the various parts of the school curriculum. Such results are particularly useful at those points where certain choices in the course of study must be made. Reference to this class list reveals that even within a small group there are pupils whose results in the quantitative or numerical abilities, as measured by this test, are in considerable contrast to their records on the linguistic parts. Frank Nuncio, for instance, exceeds the results for 70 percent of the independent-school group in Grade 11 on the quantitative section of the test. In L-score, however, he is just below the independent-school median. Here is evidence that Frank may meet with more success in mathematics and related subjects than in those which will demand facility with verbal and linguistic symbols. A record showing much greater promise, so far as one can judge, for the linguistic fields is that for Elizabeth Crowley. This pupil exceeds only 5 percent of the independent-school group in her grade on the quantitative part of the test but has a percentile rating of 61 in L-score. Probably her achievement in sections of the curriculum demanding the ability to handle verbal material will be higher than in those parts of the course of study relying heavily upon numerical ability.

What are some of the conclusions which one can draw from the reading test results? How would these conclusions be expected to modify the picture of academic aptitude given by the American Council test? In general, it will be seen that results on the reading test are fairly similar in standing to results on the linguistic part of the American Council test, for this group. A number of the pupils seem to be doing better

INTRODUCTION TO TESTING

on the reading test than one might predict from the academic aptitude test. Elizabeth Crowley, Gladys Gillman, and Frank Nuncio all surpass a considerably higher proportion of their independent-school colleagues on the reading test than they do on the linguistic part of the psychological examination. Alfred Diamond, on the other hand, has a linguistic score with a percentile rating of 70, a record which suggests that he should be capable of attaining much higher scores on the reading test than are illustrated in this list. There is a considerable difference between his L-score percentile rating of 70 and his total reading comprehension score, which has a percentile rating of 43. An inspection of the part scores on the reading test for this pupil suggests that corrective work for him might emphasize better skills for story comprehension and paragraph comprehension. His level on the vocabulary part is much closer to his percentile rating on the linguistic part of the psychological test.

Although the results of academic aptitude tests are very useful in the interpretation of scores on reading and other achievement tests, comparisons between the results for these two kinds of tests should be made with some caution, particularly for pupils near the top or the bottom of the distributions. In their efforts to bring achievement test percentiles up to percentiles on intelligence or academic aptitude tests, well-meaning teachers are sometimes unreasonably severe on pupils with very high academic aptitude scores. Even with tests that are rather highly correlated, as are tests of intelligence and reading ability, where the correlation is frequently of the order of about .7, the regression effect is considerable. In other words, because of a statistical phenomenon which cannot be offset by greater effort on the part of the

pupil or by any teaching procedures whatsoever, the pupils who are toward the extremes of the distribution on a test of academic aptitude may regress toward the mean on the reading test or other achievement test. For example, if a pupil obtains a percentile rank of 95 in Q-score on the American Council Psychological Examination and one of 72 on the Cooperative Plane Geometry Test, the difference may be due largely to regression effect. Teachers should use care not to interpret such differences as lack of effort or need for more intensive instruction. For pupils very high or very low in ability as measured by intelligence or academic aptitude tests, moderate differences between academic aptitude and achievement test results usually should be ignored.

At the same time, it is highly desirable to take academic aptitude test and reading test results into account when studying achievement test records. Often, academic aptitude and reading test scores are obtained in the fall, whereas achievement tests are administered in the spring. Since class lists become unwieldy if they are extended to accommodate all test scores, it becomes essential to set up an individual cumulative record in order to relate properly all test results for the individual and to accumulate them not only from fall to spring but from year to year. The next chapter is devoted to a fairly thorough discussion of the individual cumulative record.

SUMMARY

Tests results can be analyzed and interpreted intelligently only after familiarity is gained with terminology and commonly used techniques employed in the field of testing. At first these may seem complicated to a teacher having no special

HOW SHALL WE ANALYZE AND INTERPRET TEST RESULTS?

preparation in tests and measurements. However, the elementary concepts needed for understanding the test record can be mastered without difficulty. Usually it is helpful to provide in-service training sessions dealing with the technical aspects of test analysis and interpretation when the school reaches that point in the test program where the teachers are ready to use the scores of the pupils in their classes.

Statistical methods assist the teacher in ana-

lyzing class performance and in comparing the individual with the group. Published test norms, usually in the form of percentiles, grade equivalents, or age equivalents, enlarge the basis of comparison to include pupils in other schools described with regard to representativeness by the test publisher.

Various illustrations show procedures and forms which may be employed effectively in the local scoring situation.

SUGGESTIONS FOR FURTHER READING

1. Adkins, Dorothy C., *Construction and Analysis of Achievement Tests*, Washington, United States Government Printing Office, 1947, pp. 79-209.
2. Crawford, A. B., and Burnham, P. S., *Forecasting College Achievement*, New Haven, Yale University Press, 1946, pp. 21-75.
3. Flanagan, John C., *Scaled Scores*, New York, Cooperative Test Service of the American Council on Education, December, 1939.
4. Garrett, Henry E., and Woodworth, R. S., *Statistics in Psychology and Education*, New York, Longmans, Green and Company, 3rd ed., 1947.
5. Greene, Edward B., *Measurements of Human Behavior*, New York, The Odyssey Press, Inc., 1941, pp. 39-64, 675-690.
6. Greene, Harry A., Jorgensen, Albert N., and Gerberich, J. Raymond, *Measurement and Evaluation in the Secondary School*, New York, Longmans, Green and Company, 1944, pp. 574-607.
7. Lindquist, E. F., *A First Course in Statistics*, Boston, Houghton Mifflin Company, 1938.
8. Micheels, W. J., and Karnes, M. R., *Measuring Educational Achievement*, New York, McGraw-Hill Book Company, 1950, pp. 414-453.
9. Remmers, H. H., and Gage, N. L., *Educational Measurement and Evaluation*, New York, Harper & Brothers, 1943, chaps. 21-22.
10. Ross, C. C., *Measurement in Today's Schools*, New York, Prentice-Hall, Inc., 2nd ed., 1947, pp. 216-281.
11. Traxler, Arthur E., *Functions of the Educational Records Bureau in Comparable Measurement*, Educational Records Supplementary Bulletin H, New York, Educational Records Bureau, October, 1947.
12. Traxler, Arthur E., *Techniques of Guidance*, New York, Harper & Brothers, 1945, pp. 174-184.
13. Walker, Helen M., *Mathematics Essential for Elementary Statistics*, New York, Henry Holt and Company, rev. ed., 1951.
14. Weitzman, Ellis, and McNamara, Walter J., *Constructing Classroom Examinations*, Chicago, Science Research Associates, 1949, pp. 94-107.
15. Wood, Ben D., and Haefner, Ralph, *Measuring and Guiding Individual Growth*, New York, Silver Burdett Company, 1948, pp. 399-427.

How Shall We Record Test Results?

IN THE preceding chapters frequent mention has been made of the need for systematic accumulation of the results of tests. This aspect of testing and the use of test results is really basic to the process of individualized education. It would be difficult indeed to determine rate, amount, and direction of academic development without relating current test results to previous test performance. Actually, a testing program brought up to the point of interpreting and analyzing results without a plan for keeping results readily available would show little ultimate return on the investment of testing costs. An essential part of planning and carrying out a testing program, then, is the recording of test results.

It may be argued that class lists, such as those described in the preceding chapter, provide an accurate record of results which may be filed for future reference. It will be remembered that such records do show individual scores and individual percentile rankings, usually with a description of the norms used typed on the list. Even a cursory analysis of the uses to which test records may be put will reveal inefficiency in such a method of recording results permanently.

Consider, for example, the following illustration:

One secondary-school principal decided that the class lists showing results of achievement tests should be filed in the offices of the various departments where they would be readily available to teachers. Since there was frequent reference to aptitude test scores in the principal's office, he decided that the lists showing scores on aptitude tests should be kept in his own files. In a discussion of the uses of test scores with his teachers, the suggestion developed that it would be helpful to relate results of achievement tests to the aptitude scores. Following this meeting, an attempt was made to carry out the suggestion, but it proved difficult to locate the various records. Frequently the needed report could not be found. Other weaknesses in the plan became evident when the principal arranged a conference with one of the students and his parents to discuss plans for the boy's college entrance. In preparing for the conference, the principal decided to assemble all of the boy's test scores together with teachers' marks and comments and other personal data. In trying to locate the test scores, he found that the lists for English had been filed in the departmental office but the

HOW SHALL WE RECORD TEST RESULTS?

language lists were still in the hands of the individual teachers. On the Latin class list no score was shown for this student. When the Latin teacher was consulted, he remembered that the boy was ill on the day the class took the test, and that the test was given to him separately after his return to school. The principal went back to the school office and after some hunting found the absentee test report filed carefully by the secretary under "Attendance Records."

One who has had such an experience does not need to be told that all information concerning one pupil's test scores should be assembled in one place and set up on an individual basis, particularly when pupils have been tested in more than one program. Then the school administrator, the adviser, or the classroom teacher who consults the individual record sees not only the pupil's present status but also the route by which he reached it. This overall picture of individual growth presents a relatively complete pattern, not a cross-sectional view, such as that obtained from a single testing program, nor a longitudinal view of development in a single field. It is true that a record from a single testing program or a developmental picture in a single subject will be better than no information at all. On the other hand, the aim of modern education to know the pupil as an individual can be realized only if each teacher dealing with the pupil has as complete a picture as possible.

One method which has been employed for keeping a cumulative record of test scores is that of setting up individual folders and filing in the folder for each pupil the results of successive testing programs. Many test publishers provide individual report forms for the various kinds of tests. These may be printed on the

cover or back page of the test booklet, or they may be furnished as supplementary test supplies.

Typical of such forms are the individual profile sheets for the Differential Aptitude Tests, the Cooperative Achievement Tests, and the Kuder Preference Record, Vocational, illustrated in Figures 9 through 11. Forms like these are easily filed in individual folders, thus keeping in one place all test results for a given student. With the profiles for an individual pupil spread out before her, the teacher or counselor can study achievement in relationship with aptitude and interests or can consider other interrelationships which may be helpful in understanding the pupil. Also, the graphic presentation of the test data facilitates understanding of results on the part of the teacher and is an aid in explaining the record to the pupil or to his parents.

It will be helpful at this point to comment on the scores entered in the illustrated forms. On the Differential Aptitude Test profile, a bar has been drawn from the fiftieth percentile line to the student's percentile placement on each of the tests. This facilitates observation of distance above or below the median for each obtained score and is in accord with instructions supplied by the authors of the test battery. The profile shows James Crawford to be somewhat below average in spelling, sentences, and verbal abilities when compared with other boys at a similar grade level. He is above the fiftieth percentile in all of the other part scores and is particularly high in numerical, abstract reasoning, and space relations skills. The pattern of aptitude scores suggests that James may be expected to handle subjects such as mathematics and science without difficulty and that he may develop more slowly with regard to those parts of the curricu-

**INDIVIDUAL
REPORT
FORM**

DIFFERENTIAL APTITUDE TESTS

G. K. Bennett, H. G. Seashore, and A. G. Wesman

THE PSYCHOLOGICAL CORPORATION

New York 18, N. Y.

CRAWFORD, JAMES R.

M 14-3 9

NAME

SEX

AGE
YRS. MOS.

GRADE

SUMMITVILLE, OHIO

A

GRADE 9 BOYS

SEPT 14, 1950

PLACE OF TESTING

FORM

NORMS USED

DATE OF TESTING

Raw Score

Standard
Score

Percentile
99

95

90

80

75

70

60

50

40

30

25

20

10

5

30

Percentile
99

95

90

80

75

70

60

50

40

30

25

20

10

5

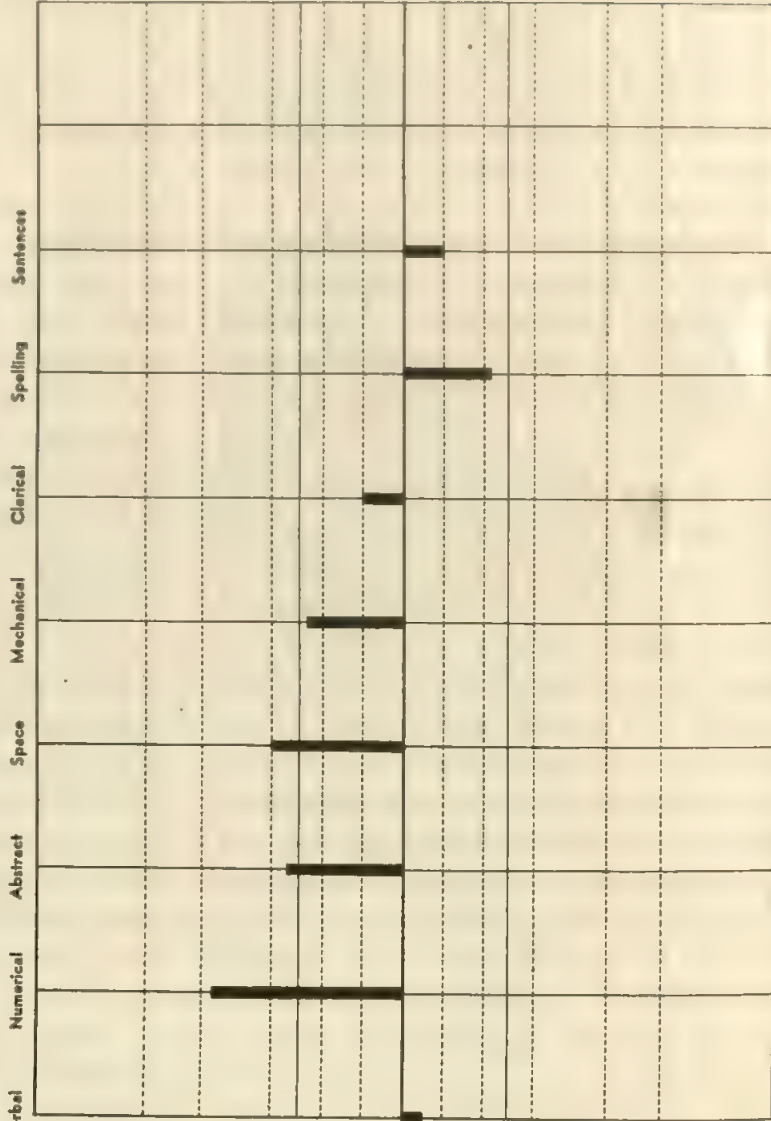


FIGURE 9. Differential Aptitude Test Profile. (Used with permission of Psychological Corporation.)

HOW SHALL WE RECORD TEST RESULTS?

lum making demands on verbal and written expression skills.

When achievement tests are administered in the spring, the results for James are generally in line with the pattern of accomplishment predicted by the aptitude scores. James's achievement in mathematics and science is well above average, while his scores on the English and social studies tests are nearer to the Grade 9 medians. Actually, the boy has attained a somewhat higher level of achievement in the two "verbal" subjects than perhaps would be expected on the basis of the aptitude profile.

The Kuder Preference Record was administered to James's class during the same week the achievement tests were given. The results of the interest test are used at the end of Grade 9 in the Summitville Junior High School as an aid in counseling concerning choice of curriculum in the senior high school. It will be seen that James's interest profile is in general agreement with aptitude and achievement results. Literary interest is low, while scientific, artistic, and mechanical interests are high. The computational interest score may seem to be somewhat out of line with the numerical aptitude and mathematics achievement displayed in the other profile. It appears that routine, computational types of work with numbers do not interest the boy. However, both numerical and abstract reasoning skills are required in scientific pursuits, the area in which James's interests are particularly high.

The high artistic interest score may bear further investigation. Specific exploration of art aptitude may disclose promise for development in this field. However, artistic interests blend well with scientific and mechanical interests in some such vocational pursuit as architecture, drafting, or perhaps machine design or mechani-

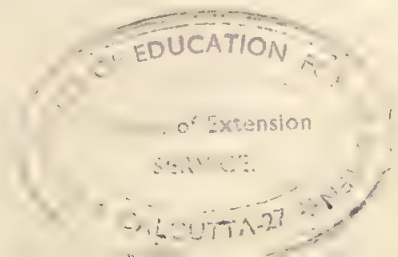
cal engineering. Furthermore, the combination of numerical, abstract reasoning, and space relations aptitudes revealed in the D.A.T. profile supplies other evidence that the areas toward which interests are directed should receive careful consideration in deciding the direction of further schooling.

It would not be advisable for James and his teacher or counselor to attempt a decision concerning kind of high-school preparation on the basis of these test scores alone. The choice might be college preparatory with an engineering degree as the future objective; or it might be perhaps the vocational curriculum either as terminal schooling or as preparation for junior college or further trade or technical training. The decision should be made on the basis of all available information which may have bearing on the choice—the previous school record, other test scores, attitude of the parents toward further schooling, financial limitations, study and work habits, and so on.

However, this illustration suggests the value of bringing individual test scores together and points out some of the uses of typical individual record forms supplied by test publishers.

Although the method of accumulating test records by continuous filing of individual report forms may suit local needs in some situations, it has limitations which should be pointed out.

First of all, if the testing program is fairly comprehensive so that several tests are administered each year, the file will become quite bulky. This difficulty is increased by the usual practice of adding other personal data to the file, such as letters from parents, anecdotal records, and listing of course grades and marks. Not only does this impose a problem with regard to filing space, but it also may create confusion in using



INDIVIDUAL PROFILE FOR COOPERATIVE ACHIEVEMENT TEST SCORES

Junior High School Form

Name CRAWFORD, JAMES P. Date of Birth JUNE 10, 1936 Sex M

Class of 19 50 School SUMMITVILLE JUNIOR HIGH City SUMMITVILLE, OHIO

ENGLISH							SOCIAL STUDIES 7, 8, 9				SCIENCE 7, 8, 9				MATHEMATICS 7, 8, 9					GENERAL SCIENCE		ELEMENTARY ALGEBRA	Sealed Scores	
Expression	Reading	Comprehension	Interpretation	Application	Analysis	English TOTAL	Sealed Scores Per English	Sealed Scores	FACTS, SKILLS & APPLICATIONS	TERMS & CONCEPTS	COMPREHENSION & INTERPRETATION	TOTAL	FACTS, SKILLS & APPLICATIONS	TERMS & CONCEPTS	COMPREHENSION & INTERPRETATION	TOTAL	SKILLS	FACTS, TERMS, & CONCEPTS	APPLICATIONS	APPRECIATION	TOTAL	GENERAL SCIENCE		ELEMENTARY ALGEBRA
75							75	90																90
74							74	89																89
73							73	88																88
72							72	87																87
71							71	86																86
70							70	85																85
69							69	84																84
68							68	83																83
67							67	82																82
66							66	81																81
65							65	80																80
64							64	79																79
63							63	78																78
62							62	77																77
61							61	76																76
60							60	75																75
59							59	74																74
58							58	73																73
57							57	72																72
56							56	71																71
55							55	70																70
54							54	69																69
53							53	68																68
52							52	67																67
51							51	66																66
50							50	65																65
49							49	64																64
48							48	63																63
47							47	62																62
46							46	61																61
45							45	60																60
44							44	59																59
43							43	58																58
42							42	57																57
41							41	56																56
40							40	55																55
39							39	54																54
38							38	53																53
37							37	52																52
36							36	51																51
35							35	50																50
34							34	49																49
33							33	48																48
32							32	47																47
31							31	46																46
30							30	45																45
29							29	44																44
28							28	43																43
27							27	42																42
26							26	41																41
25							25	40																40
24							24	39																39
23							23	38																38
22							22	37																37
21							21	36																36
20							20	35																35
19							19	34																34
18							18	33																33
17							17	32																32
16							16	31																31
15							15	30																30
14							14	29																29
13							13	28																28
12							12	27																27
11							11	26																26
10							10	25																25
9							9	24																24
8							8	23																23
7							7	22																22
6							6	21																21
5							5	20																20

Notes and Interpretations: TESTS GIVEN MAY 16-18, 1951: ENGLISH, FORM T;
SOCIAL STUDIES, FORM Y; SCIENCE, FORM Y; MATHEMATICS, FORM Y

Directions for filling out this Chart are given on the back of the sheet.
 Norms for each grade are indicated at the left of each scale as follows: seventh grade, ⑦; eighth grade, ⑧; ninth grade, ⑨.
 In the case of the General Science and Elementary Algebra Tests, the ⑨ indicates the norm for groups with two semesters of study at the end of the ninth grade.

FIGURE 10. Individual Profile for Cooperative Achievement Tests. (Copyright 1950 by the Educational Testing Service. Used with permission of Educational Testing Service, Princeton, New Jersey.)

Revised August, 1949

REPRODUCTION BY ANY MEANS STRICTLY PROHIBITED

**SELF-INTERPRETING
PROFILE SHEET**
for the
**KUDER PREFERENCE RECORD
VOCATIONAL**

Form C

BOYS and GIRLS

DIRECTIONS FOR PROFILING

1. Copy the V-Score from the back page of your answer pad in the box at the right.

If your V-Score is 37 or less, there is some reason for doubting the value of your answers, and your other scores may not be very accurate. If your V-Score is 45 or more, you may not have understood the directions, since 44 is the highest possible score. If your score is not between 38 and 44, inclusive, you should see your adviser. He will probably recommend that you read the directions again, and then that you fill out the blank a second time, being careful to follow the directions exactly and to give sincere replies.

If your V-Score is between 38 and 44, inclusive, go ahead with the following directions.

2. Copy the scores 0 through 9 in the spaces at the top of the profile chart. Under "OUTDOOR" find the number which is the same as the score at the top. Use the numbers under M if you are a boy and the numbers under F if you are a girl. Draw a line through this number from one side to the other of the entire column under OUTDOOR. Do the same thing for the scores at the top of each of the other columns. If a score is larger than any number in the column, draw a line across the top of the column; if it is smaller, draw a line across the bottom.

3. With your pencil blacken the entire space between the lines you have drawn and the bottom of the chart. The result is your profile for the Kuder Preference Record—Vocational.

An interpretation of the scores will be found on the other side.

Published by SCIENCE RESEARCH ASSOCIATES,
228 South Wabash Avenue, Chicago 4, Illinois.
Copyright 1948, by G. Frederic Kuder.

Copyright under International Copyright Union.

All rights reserved under Fourth International American Convention (1910).
Printed in the U.S.A. Copyright 1948 in Canada.

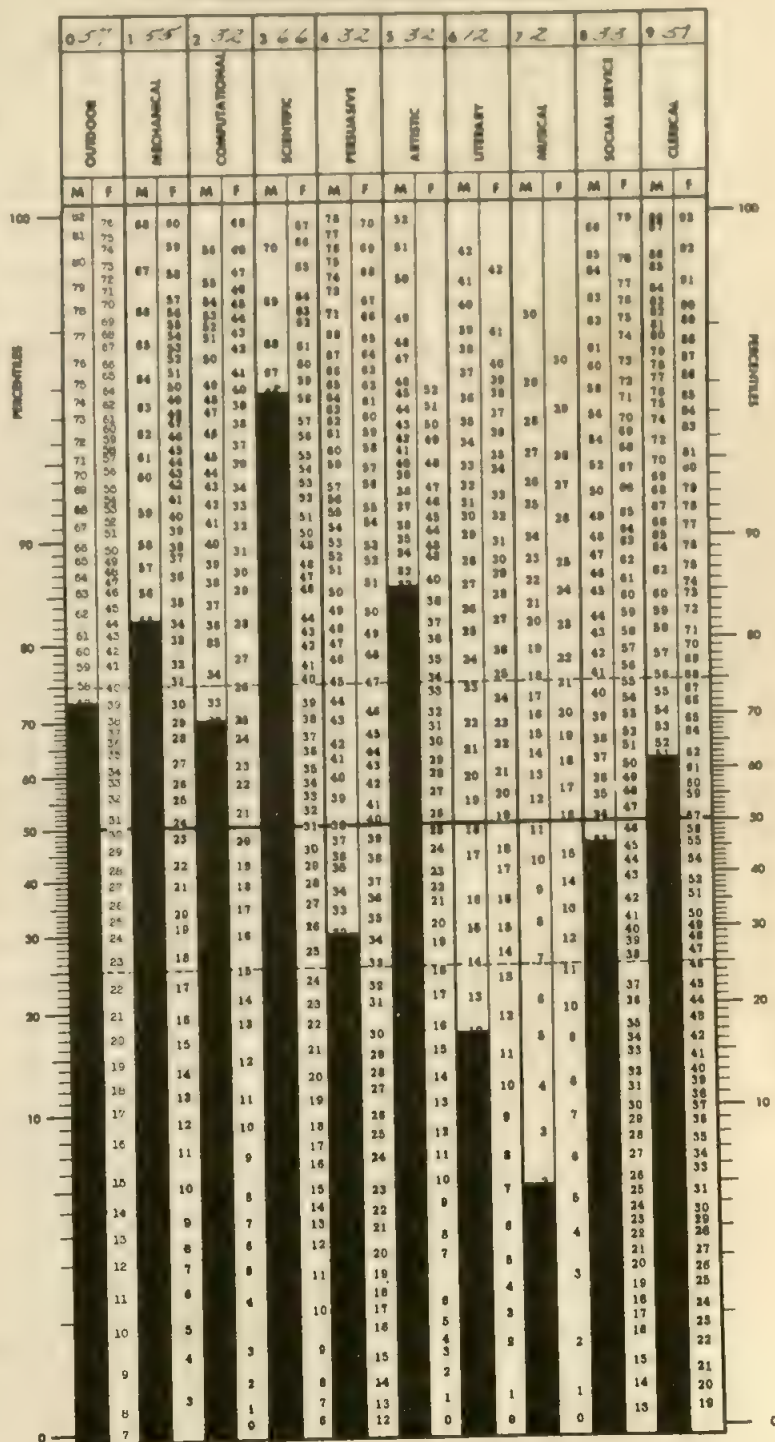


FIGURE 11. Kuder Preference Record Profile Sheet. (Used with permission of Science Research Associates.)

INTRODUCTION TO TESTING

the file contents. Over a period of a few years the individual folder may become virtually a jigsaw puzzle as one attempts to piece together odd bits of information which may have bearing on a particular problem or upon a particular decision.

A second limitation quite important from the standpoint of using test results is that the individual file system may focus attention on a cross-sectional view of the pupil rather than on the developmental aspect of academic growth. In the illustrated record, for example, although the results of one kind of test are related to those of other kinds, the profiles do not show anything with regard to past accomplishment. The three profiles show simply a picture of present status with no clues as to rate or direction of previous development.

This limitation does not apply to all individual profile forms. For example, the individual form for Cooperative achievement test scores illustrated in Figure 10 provides for an accumulation of scores over a three-year period by the use of different symbols in making profile entries for the three grades. Thus, if achievement were measured in Grades 7, 8, and 9 by use of different forms of the same tests (the Cooperative tests in lower-level English and in social studies, science, and mathematics for Grades 7, 8, and 9), and the scores were entered individually on this profile form, the teacher could observe the extent of individual growth by relating present and past performance on each of the tests.

Unfortunately, not all profile forms lend themselves easily to such cumulation of results. The adequacy of such forms as the Cooperative tests profile is limited, too, by the fact that different kinds of tests are used at different levels of study, and the period of growth (the grade range) covered by any one is usually relatively

small. So, although some profile forms provide for limited accumulation of individual test data, in the main the focus of attention is upon results obtained at a given time.

The value of a single cumulative record card becomes apparent. Although it is usually not possible to construct a single record form which will provide for entering all the data related to a particular individual case, many satisfactory record forms have been devised which provide organization and system to the gathering of data regarding the important areas of needed information. Large schools or school systems frequently set up their own cards in order to have records adapted to their own testing program and their particular form of organization for guidance services. Some such forms provide a great deal of space for records of the home-room teacher or the pupil's adviser, school attendance, and related data. Others devote most of their space to school grades. In planning a new record system it is wise for the entire faculty or a faculty committee to outline the school's needs and then select a type of card adapted to these needs or else devise such a record form.¹

It may be appropriate to suggest here certain guiding principles for the selection or construction of a cumulative record form.

1. It should agree with the objectives of the local school.
2. It should be the result of the group thinking of a faculty committee.
3. It should either provide for a continuous record of the development of the pupil from the

¹ A school which wants to adopt a form for general use may be interested in looking at a sample set of records collected from public and independent schools and colleges. Such sets of records are available on loan from the Educational Records Bureau. There is a fee of \$2.00 for this service to cover costs of mailing, checking, and keeping the sets up to date.

HOW SHALL WE RECORD TEST RESULTS?

first grade to the end of the junior college or be one of a series of forms which make provision for such a record.

4. It should be organized by time sequence; that is, it should be set up by yearly divisions which run throughout the form.

5. It should contain ample and carefully planned space for a record of the results of all types of tests and for an explanation of the norms in terms of which the results were interpreted.

6. It should provide for the annual recording of personality ratings or behavior descriptions which represent the consensus of the pupil's counselors and teachers.

7. While it should be as comprehensive as possible, it should be simple enough to avoid overwhelming the clerical resources of the school.

8. It should be accessible to the teachers as well as to the counselors and principal. Highly confidential information which the counselor may possess should be filed elsewhere.

9. The record form should be reevaluated periodically and revised as needed to take account of educational change and progress.

While it is recognized that the need for coordinating all types of personal data should be emphasized, this discussion will be particularly concerned with item 5, dealing with recording of test results. Examination of various record forms reveals a variety of methods employed in keeping the cumulative record of test scores. Some cards simply provide spaces for entering test data with columnar headings such as *Date of Test*, *Name of Test*, *Raw Score*, and *Percentile*. Others use this method with some such space division as *Achievement Test*, *Mental Test*, *Special Aptitude Test*, and so on. Unless a column is added for entering a description of

the norm group used, it becomes necessary to employ symbols or footnotes to interpret adequately percentile ranking for the variety of test results which may be entered on one form. It is necessary at all times to identify and describe carefully on the cumulative record the norm base from which percentiles for various tests are drawn. Otherwise, unwarranted comparisons may be made in considering cumulative entries.

The cumulative picture of growth is presented more clearly if a section of the card is devoted to graphic presentation of test results. By proper use of symbols and legends, it is possible frequently to extend the individual profile over a period of several years. This procedure receives some emphasis in illustrative material shown in later portions of this chapter.

A great deal could be written about the value and uses of this cumulative method of recording test scores. Tests are not perfectly reliable nor is human nature perfectly stable. It is entirely possible for the results of a single test to misrepresent seriously the aptitude or achievement of the pupil, since one test usually samples the activities or materials of only one subject field and since the results may be influenced by the health or state of mind of the pupil. Too, academic growth is not a steady, continuous process. There are spurts as well as periods of relatively slow progress. A single test score provides a clue as to status at a given point in the growth pattern, and if this should be at an unusual period of maturation, false impressions may result. Although the results of any one test may be seriously in error, it is not probable that all tests taken over a period of years will err in the same direction, and thus the general picture will be fairly accurate.

Probably it will be helpful to present at this

INTRODUCTION TO TESTING

point samples of long-term records which illustrate the fact that test results do definitely tend to show a significant trend when accumulated over a period of time. To report authentic data, the files and records of the Educational Records Bureau have been employed in developing the following illustrations. Because the member schools of the Bureau are made up mostly of independent or private schools, the illustrations are based on scores of independent-school pupils and are interpreted in terms of independent-school norms. Nonetheless, the value and usefulness of long-term cumulative records can be pointed out, even though the scores reported and the norms used are somewhat higher than those usually derived from a public-school situation.

One record form employed by Bureau member schools is designed both for a cumulative record card in the elementary grades and as an admissions record for the secondary school. The form was devised by the Bureau's Subcommittee on Relations Between Elementary and Secondary Schools in coöperation with the Secondary Education Board. The history of the cumulative record card, the procedure used in developing this one, and the uses of the card were summarized in an article published several years ago in the *Elementary School Journal*.² In an effort to make the card useful for transfer purposes, the admissions forms of about sixty secondary schools were collected and a tabulation of items of information called for was made. The current revision of the form reproduced on pages 78, 79, and 80 incorporated suggestions made by elementary schools holding membership in the Educational Records Bureau and by

elementary-school members of the Secondary Education Board.

The main form consists of a card providing space for recording information about the student's school activities, general health, and personal characteristics. The back of this card provides for a record of school marks and achievement test scores. A second card, which may be used to supplement this record, provides space for a graphic representation of test scores in terms of percentiles.

The cumulative record for Albert Stanley, illustrated in Figures 12, 13, and 14, shows that the Glenwood School judged that he came from a good home; he was regular in attendance, participated in a variety of activities, and had normal health and a good personality. He was a little older than the average independent-school boy for his grade. This fact is explained when one notices that he lost some time in the earlier grades as his family moved from one place to another.

His achievement, as judged by both school marks and test scores, is consistently close to or above average, in line with his academic aptitude, which seems to be, on the whole, above that of about 60 percent of the independent-school group. The graphic form in Figure 14 shows that most of his test scores were above the medians for independent-school pupils in corresponding grades. With a form of this sort, where achievement is graphed in terms of percentile ratings, a pupil who progresses at a normal rate for the group with which he is being compared will tend to maintain about the same percentile ratings from year to year. Albert's achievement in reading, as measured by the Traxler Silent Reading Test given in the fall of the sixth and seventh grades, is just below the independent-school median. He is just above the

² Arthur E. Traxler, "A Cumulative Record Card for the Elementary School," *Elementary School Journal*, 40:45-54 (September, 1939).

HOW SHALL WE RECORD TEST RESULTS?

median in reading by the time he reaches the eighth grade. In the spring of his eighth-grade year, however, he exceeds the results for only one-third of the independent-school group on the reading comprehension section of the Co-operative English Test. From this part of his record one would judge that a little extra attention to his reading skill might help him bring some of his class work up to a point more nearly equaling his academic aptitude.

The fact that he was a little below the independent-school norm in the literature section of the Metropolitan Achievement Test, which he took in the spring of the seventh grade, and that he is above only about one-third of the independent-school group beginning Latin in the eighth grade, is consistent with this picture of a pupil who is a little lower in the linguistic area than he is in certain other parts of the curriculum. It will be noted that he has been consistently high in the social studies and has done well in arithmetic as measured by the appropriate sections of the general achievement tests. In the fall of the eighth-grade year he made a good record on the Reavis-Breslich Diagnostic Test in Arithmetic. His results on the Co-operative Mathematics Test for Grades 7, 8, and 9 are considerably lower. This may, of course, reflect a change in the type of mathematics work which he has been doing in the eighth grade. From the record of school marks, we find that in the second term in Grade 8 he began the study of algebra, and he does not seem to be doing quite so well in this subject as he did in arithmetic courses taken in previous terms.

It is apparent that Albert's record tends to be average or superior in terms of the independent-school norms. It is also evident that the information obtained on the form should prove a good basis for appraisal, placement, and guidance by

his present school or by any school which he may enter.

Another cumulative record card which has been used widely by secondary-school members of the Bureau as well as by some of the elementary schools is an adaptation of the American Council on Education Cumulative Record Card. This record provides, on the front, space for school marks and test scores to be recorded in both tabular and graphic form. On the other side of the card there is space for comments on personal characteristics and information about home background, extracurricular activities and interests, and outstanding accomplishments. A sample record shown on this card is given in Figures 15 and 16.

John Caldwell, whose test record is shown on the card, attended the Middle Ranch School in Grades 7 and 8. This school sent his record card with him to the Essex Preparatory School, where he is now completing Grade 12. We notice that this boy gives evidence of exceptionally high academic aptitude in the seventh and eighth grades. His achievement in most subjects is good, though his low literature scores agree with his adviser's comment that he does not seem to be much interested in reading. His history and civics score in the seventh grade was considerably below the independent-school median, but he makes a much better showing in Grade 8. His achievement in geography, as measured by the Metropolitan test, is lower in rank in the eighth grade than in the seventh. Changes in standing in the content subjects, of course, sometimes reflect differences between objectives of the school course and the material covered in the achievement test.

His marks in general mathematics in the lower grades are all high. The diagnostic picture of his academic aptitude, as given by his American

PREPARING FOR

BIRTHDATE MONTH DAY YEAR

MIDDLE NAME

FIRST NAME

LAST NAME

HALLAM

FEB 13, 1935

STANLEY ALBERT

1948-1949

1947-1948

1946-1947

8

7

6

SCHOOL CLENNWOOD

SUBJECT

SUBJECT

SUBJECT

ENG. I

ENG. C

ENG. D

SpELL A

SpELL B

SpELL C

FRENCH

FRENCH

FRENCH

LATIN

LATIN

LATIN

MATH. B

MATH. B

MATH. D

Geog. C

Geog. B

Geog. B

HIST. C

HIST. B

HIST. B

SCI. A

SCI. A

SCI. D

MUSIC B

MUSIC A

MUSIC C

ART

ART

ART

RANK

RANK

RANK

TEST

TEST

TEST

DATE

DATE

DATE

SCORE

SCORE

SCORE

C.A.

C.A.

C.A.

I.O.

I.O.

I.O.

P.S.

P.S.

P.S.

GRADE

GRADE

GRADE

%ILE

%ILE

%ILE

TIME

TIME

TIME

SUBJECT

SUBJECT

SUBJECT

ENG.

ENG.

ENG.

FRENCH

FRENCH

FRENCH

LATIN

LATIN

LATIN

MATH.

MATH.

MATH.

Geog.

Geog.

Geog.

HIST.

HIST.

HIST.

SCI.

SCI.

SCI.

MUSIC

MUSIC

MUSIC

ART

ART

ART

RANK

RANK

RANK

TEST

TEST

TEST

DATE

DATE

DATE

SCORE

SCORE

SCORE

C.A.

C.A.

C.A.

I.O.

I.O.

I.O.

P.S.

P.S.

P.S.

GRADE

GRADE

GRADE

%ILE

%ILE

%ILE

TIME

TIME

TIME

SUBJECT

SUBJECT

SUBJECT

ENG.

ENG.

ENG.

FRENCH

FRENCH

FRENCH

LATIN

LATIN

LATIN

MATH.

MATH.

MATH.

Geog.

Geog.

Geog.

HIST.

HIST.

HIST.

SCI.

SCI.

SCI.

MUSIC

MUSIC

MUSIC

ART

ART

ART

RANK

RANK

RANK

TEST

TEST

TEST

DATE

DATE

DATE

SCORE

SCORE

SCORE

C.A.

C.A.

C.A.

I.O.

I.O.

I.O.

P.S.

P.S.

P.S.

GRADE

GRADE

GRADE

%ILE

%ILE

%ILE

TIME

TIME

TIME

SUBJECT

SUBJECT

SUBJECT

ENG.

ENG.

ENG.

FRENCH

FRENCH

FRENCH

LATIN

LATIN

LATIN

MATH.

MATH.

MATH.

Geog.

Geog.

Geog.

HIST.

HIST.

HIST.

SCI.

SCI.

SCI.

MUSIC

MUSIC

MUSIC

ART

ART

ART

RANK

RANK

RANK

TEST

TEST

TEST

DATE

DATE

DATE

SCORE

SCORE

SCORE

C.A.

C.A.

C.A.

I.O.

I.O.

I.O.

P.S.

P.S.

P.S.

GRADE

GRADE

GRADE

%ILE

%ILE

%ILE

TIME

TIME

TIME

SUBJECT

SUBJECT

SUBJECT

ENG.

ENG.

ENG.

FRENCH

FRENCH

FRENCH

LATIN

LATIN

LATIN

MATH.

MATH.

MATH.

Geog.

Geog.

Geog.

HIST.

HIST.

HIST.

SCI.

SCI.

SCI.

MUSIC

MUSIC

MUSIC

ART

ART

ART

RANK

RANK

RANK

TEST

TEST

TEST

DATE

DATE

DATE

SCORE

SCORE

SCORE

C.A.

C.A.

C.A.

I.O.

I.O.

I.O.

P.S.

P.S.

P.S.

GRADE

GRADE

GRADE

%ILE

%ILE

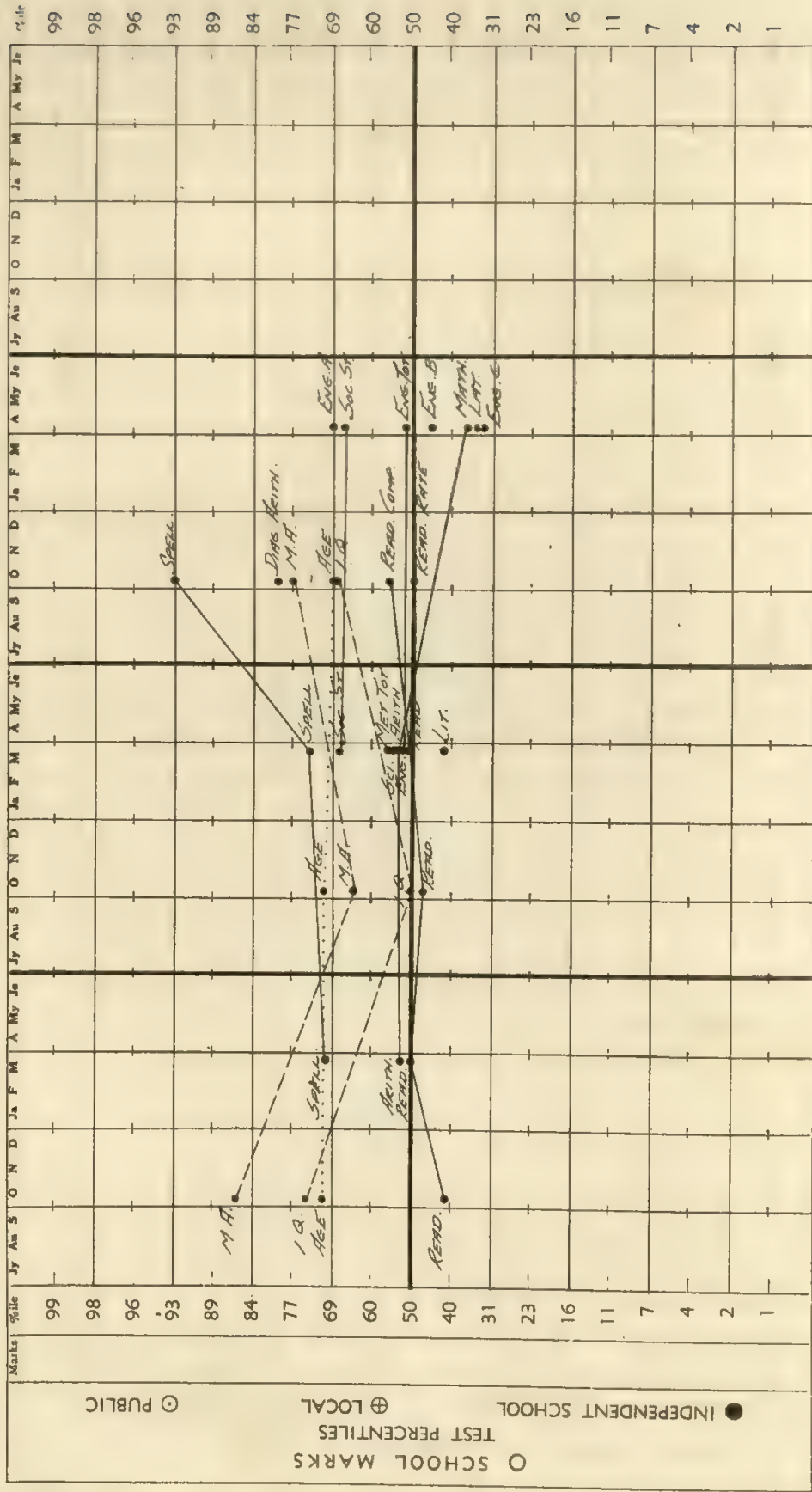
%ILE

NAME <u>STANLEY, ALBERT</u>		BIRTHDATE <u>FEB 13 1935</u>		BIRTHPLACE		HOLDER OF SCHOLARSHIP <u>NO</u>	
FATHER <u>STANLEY, CHARLES</u>		HOME ADDRESS <u>GREENWOOD, MASS</u>		BUSINESS ADDRESS		OCCUPATION <u>STATIONER</u>	
MOTHER <u>STANLEY, MARTHA B</u>		ADDRESS <u>SAME</u>				SEPARATED OR DIVORCED <u>NO</u>	
SCHOOL <u>GREENWOOD</u>		DATE OF ENTRANCE <u>SEPT 1946</u>		GRADE ENTERED <u>6</u>		PREVIOUS SCHOOL <u>WOOD HART SCHOOL</u>	

YEAR AND AGE	ADVISER	ATTENDANCE	DATE OF ENTRANCE	GRADE ENTERED	COMMENT	COMMENT	COMMENT
11	Miss Hart	REGULAR		12	Mr Jones	REGULAR	
PARTICIPATION AND ACHIEVEMENT IN ACTIVITIES							
OFFICES HELD	JUNIOR RED CROSS, CHAIRMAN FOR PICNIC			PRESIDENT SCIENCE CLUB CLASS HISTORIAN			
USE OF LEISURE TIME	INTERESTED IN NATURE STUDY, COLLECTS MOths			SCIENCE CLUB, RAISED SQUIRRELS AND RABBITS			
HOBBIES, ETC.	LITTLE INTEREST			NOT FOR HIM			
MUSIC	SHOWS			SERIAL INTEREST IN BUILDING CRABS AND FISH			
CRAFTS							
SCHOOL PAPER							
DRAMATICS	JUNIOR MASKERS						
ATHLETICS	JUNIOR SOCCER			BASEBALL SKATING CHAMPION HOCKEY			
SUMMER ACTIVITIES AND TRAVEL	VISITED IN FOUR SCHOOLS, MAPLE TRANSFERRED BY U.S. COAST AND GEOMETRIC SURVEY						
PERSONAL QUALITIES							
GENERAL HEALTH (HEARING) (LIMITATIONS)	MO. OCT. HT. 59 WT. 110 NORMAL			MO. OCT. HT. 64 WT. 120 NORMAL			
RESPONSIBILITY	FAIR			GOOD			
COOPERATIVENESS	GOOD			GOOD			
EMOTIONAL ADJUSTMENT	SEEMS GOOD, A LITTLE QUIET			EMERGING OUT OF SHELL			
SOCIAL ADJUSTMENT	GETTING ON BETTER BY 3-400			BECOMING A LEADER			
INFLUENCE	FAIR, READING SLOW			GOOD			
WORK HABITS				GOOD IF INTERESTED			
GENERAL HABITS (SEE MANUAL)							
COMMENTS AT TIME OF TRANSFER							
RESERVED FOR SECONDARY SCHOOL							

FIGURE 13. Cumulative Record Card for Independent Elementary Schools and Admission Form for Secondary Schools (Back of Card 1).

NAME	STANLEY ALBERT		BIRTHDATE	FEBRUARY 13, 1935	
YEAR	1946-1947			1947-1948	
SCHOOL	GLENWOOD				
GRADE	6			8	



NOTES

FIGURE 14. Cumulative Graphic Record of Test Percentiles and School Marks.

HOW SHALL WE RECORD TEST RESULTS?

Council Psychological Examination scores for Grades 9 through 12, provides a very helpful explanation for certain discrepancies in his test records. The high arithmetic test scores in the lower grades evidently result from a markedly superior ability in the quantitative area. He has percentile ratings high in the 90's on the quantitative sections of the American Council test throughout his high-school years. His L-scores, which are not far from the independent-school medians for the secondary grades, are in rather sharp contrast with the Q-scores.

The information on the back of the card (Fig. 16) may help us understand why John's L-score is no higher than it seems to be. His adviser in the ninth grade notes that he reads and talks slowly, though he covers textbook materials with great care. Since the speed factor is important in aptitude tests such as the American Council Psychological Examination, it is possible that a true measure of John's linguistic ability would rank him somewhat higher with respect to the independent-school pupils. One is confirmed in this impression since he has a consistently good record in French and brings his Latin results up to a high level by the second year of study.

Other comments by his advisers show a consistent picture of great ability and interest in scientific subjects. He leads his class in mathematics in the eleventh and twelfth grades, and his great interest in chemistry led to a special course in Grade 12. In this grade, only one percent of the independent-school pupils had scores as high as or higher than his in the quantitative part of the American Council test.

Both the cumulative record card for elementary schools and that for secondary schools are set up on the basis of chronological sequence horizontally maintained. The higher points on

an individual's graph of test results indicate higher percentile ratings in terms of the independent-school group at the same grade level or year of preparation in the particular subject graphed. Progress wherein the pupil maintains about the same percentile standing from year to year is indicated when the lines of the graph proceed more or less horizontally across a page. As illustrations of typical pictures found in long-term records of test scores, Figures 17, 18, and 19 are composite cards showing actual test results of pupils tested over a period of eight years.

Figure 17 shows the record for Gerald Pitt, who attended Monmouth Country Day School in Grades 5 through 7 and went on to Presbyterian Academy for Grades 8 through 12. This pupil's academic aptitude record is fairly high throughout, though in the high-school grades he shows markedly higher ability in the quantitative areas than in the linguistic. The relative preference which he shows for subjects depending on mathematical ability actually makes its appearance as early as Grade 6, where he begins to emerge as a pupil doing good work in arithmetic. It may be a matter for some surprise that he maintains his high averages in French and Latin in high school, while his mathematics test scores never quite measure up to the high quantitative scores on the American Council test. The striking improvement in this boy's spelling percentile from October testing when he was in Grade 6 to March testing at the seventh-grade level probably reflects special help in spelling in Grades 6 and 7. The moderate decline in percentile rank in this subject during the following year corresponds with the usual result when corrective instruction is discontinued.

Lydia Smith, whose results are shown in Figure 18, shows a rather different pattern, with higher linguistic abilities and somewhat lower

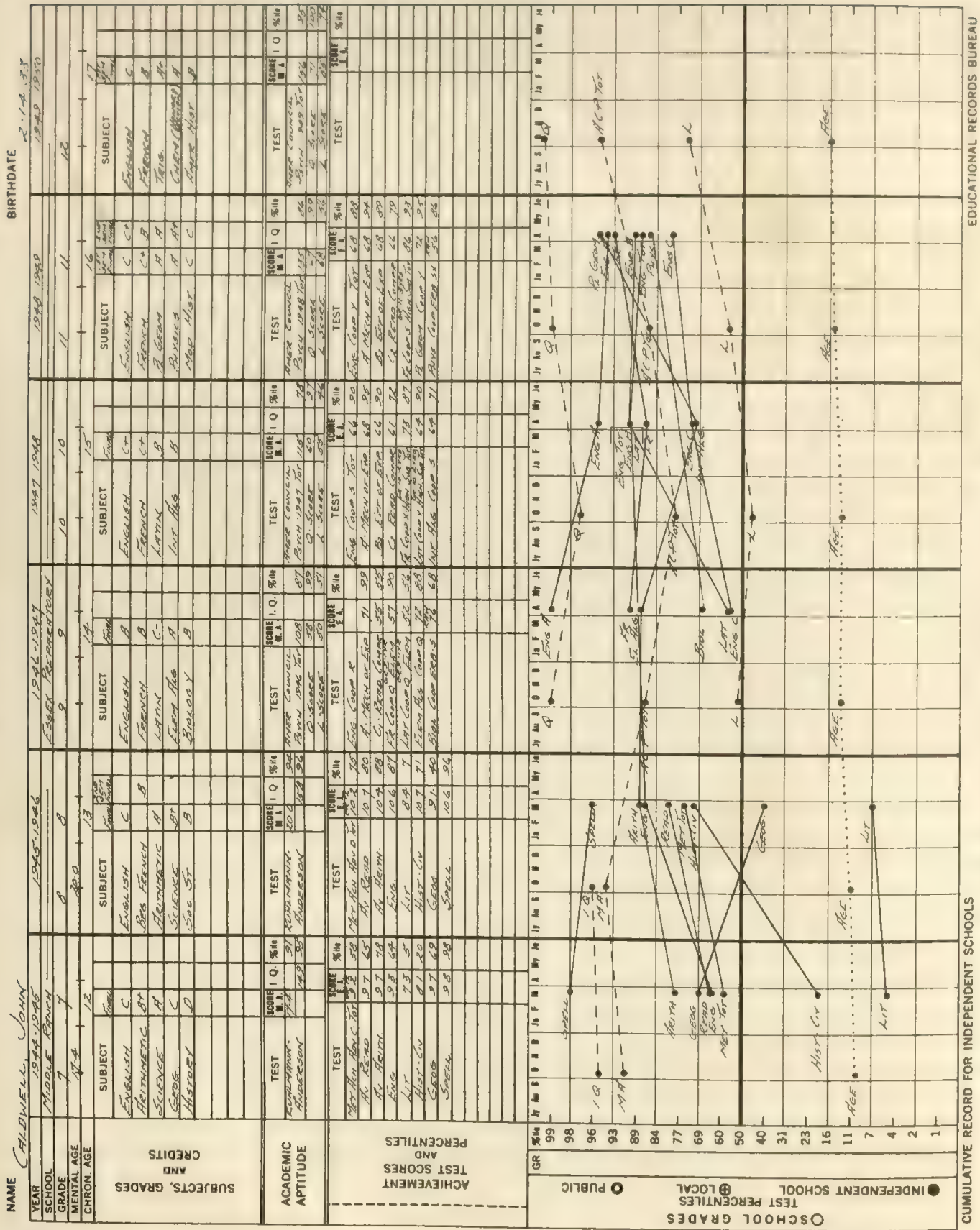


Figure 15. Cumulative Record for John Caldwell Recorded on Independent-School Adaptation of American Council on Education Cumulative Record Card.

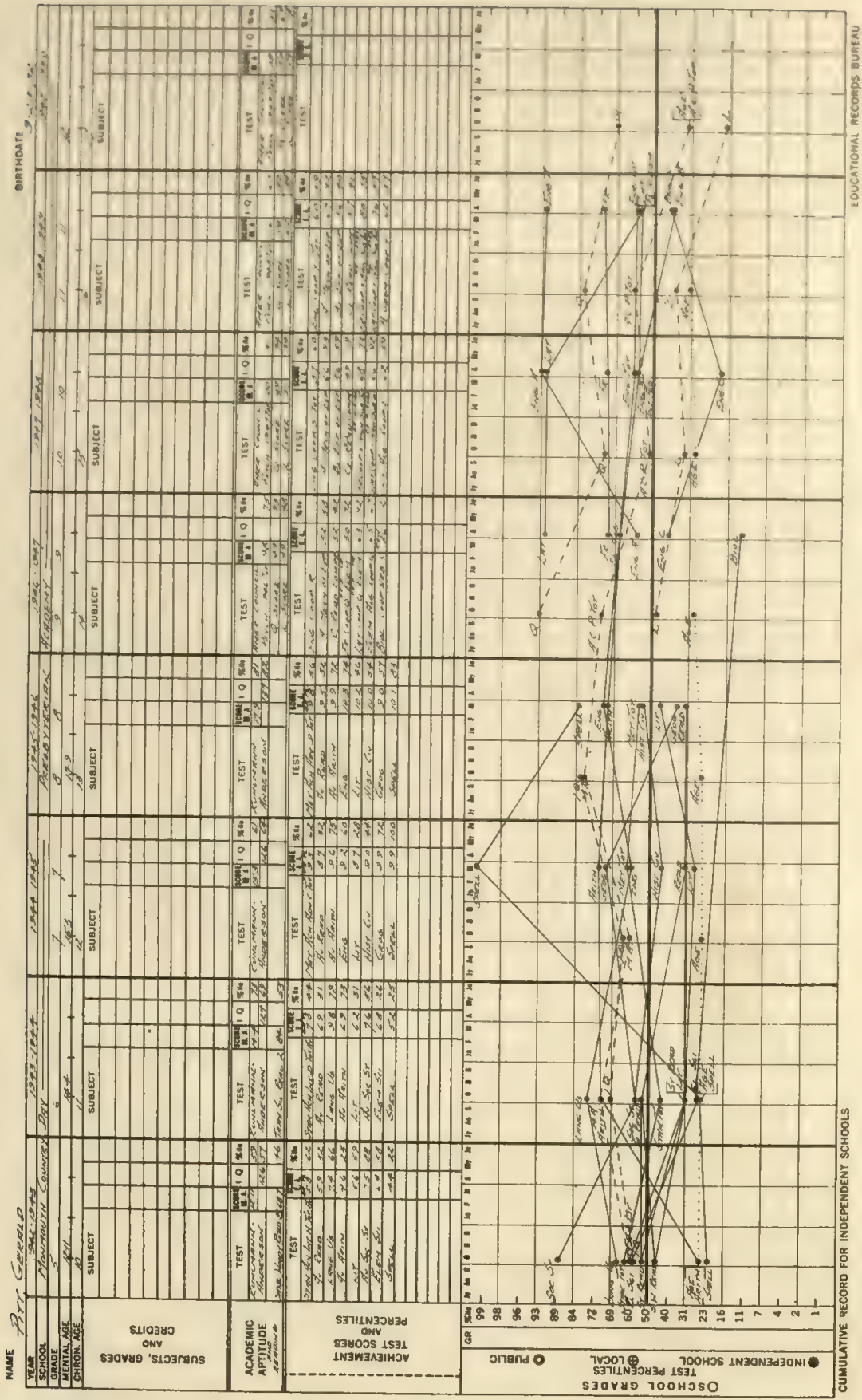


FIGURE 17. Composite Cumulative Record Card Showing Test Results of Gerald Pitt Tested over a Period of Eight Years.

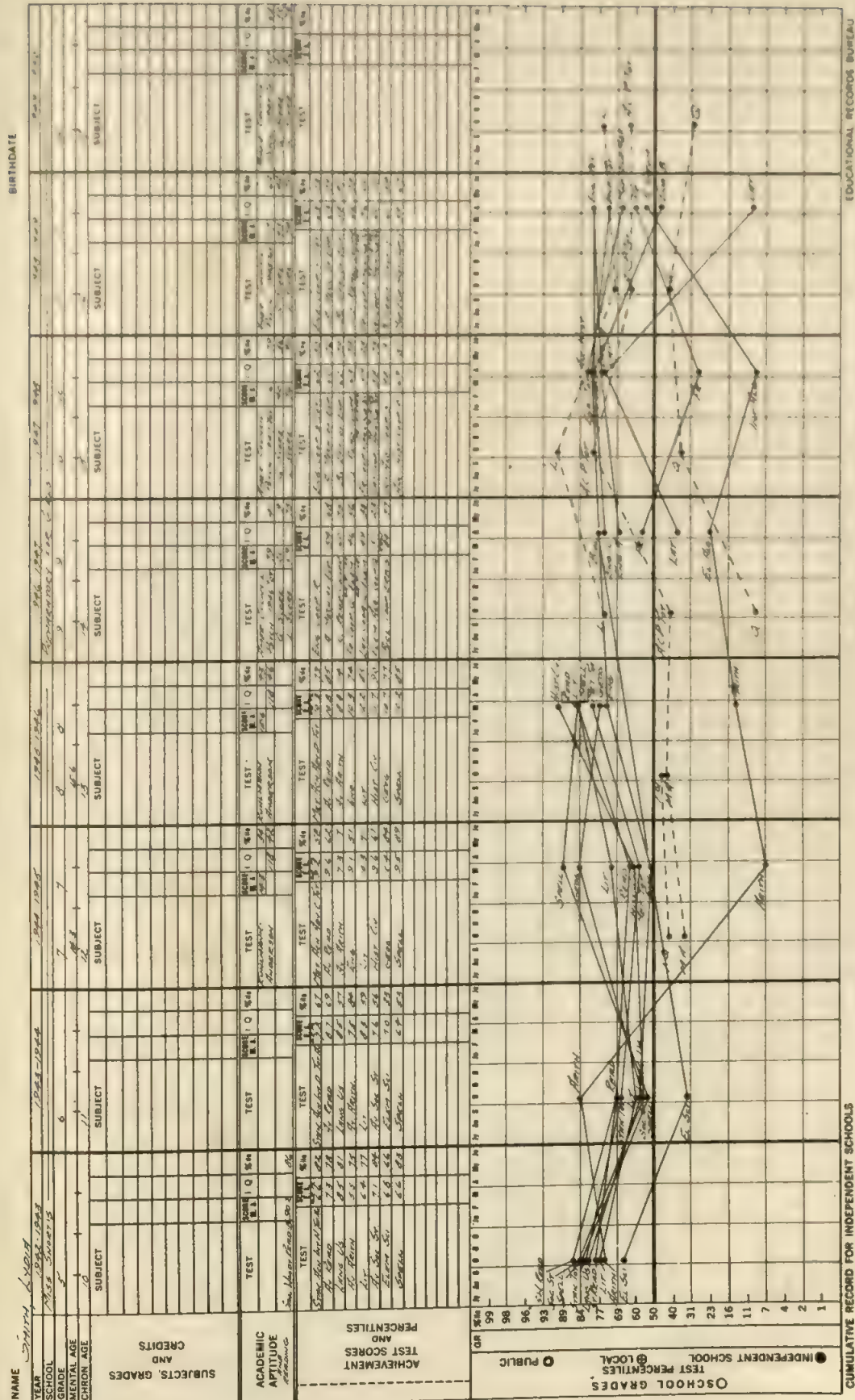


FIGURE 18. Composite Cumulative Record Card Showing Test Results of Lydia Smith Tested over a Period of Eight Years.

NAME *John Jacob*

BIRTHDATE *1903-1904*

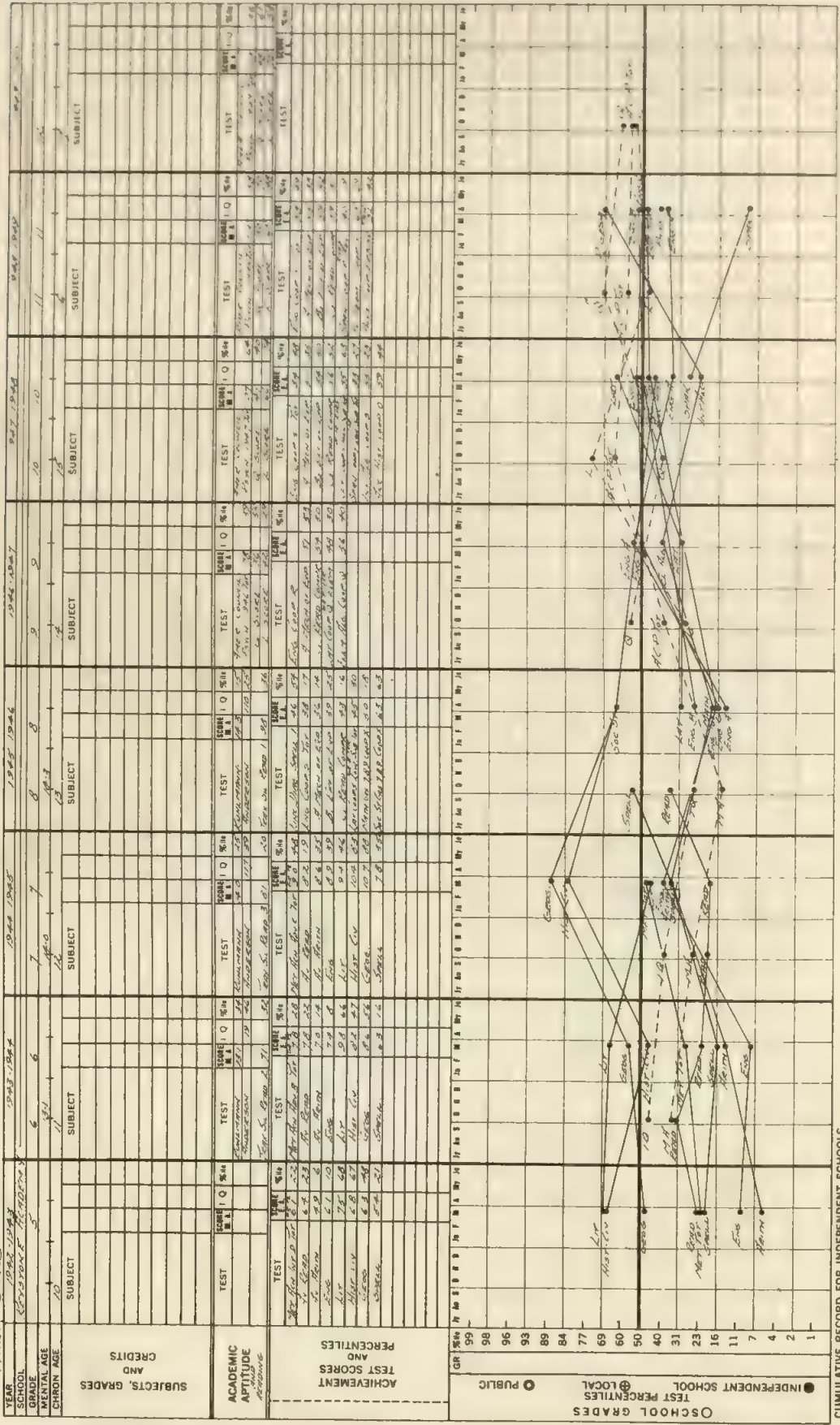


Figure 19. Composite Cumulative Record Card Showing Test Results of Jacob Hart Tested over a Period of Eight Years.

EDUCATIONAL RECORDS BUREAU

CUMULATIVE RECORD FOR INDEPENDENT SCHOOLS

HOW SHALL WE RECORD TEST RESULTS?

results on the quantitative part of the American Council test. Most of the achievement test records for this pupil are well above the independent-school medians. Certain individual subjects in the high-school curriculum, however, have evidently given her trouble. Test scores on the languages seem to be somewhat erratic.

The third record in this group, that for Jacob Hart, Figure 19, shows test scores consistently a little lower in terms of the independent-school norms than the results for the other two pupils. In the high-school years, however, he seems to be doing fairly satisfactory work in most of his subjects. The general level of his achievement test results agrees rather well with his results on the psychological examination.

The foregoing illustrations of cumulative records provide comprehensive information on the academic aptitude and achievement of individual pupils over a period of years. Results of tests are accumulated from year to year and are related to other items of personal data entered on the record form. A *whole* picture of the growing, developing boy or girl is presented so that emergent trends assume vital meaning as the teacher attempts to know and understand the child in his present situation.

SUMMARY

The general purpose of this chapter has been to discuss methods of recording test results so

that the information obtained in comprehensive annual testing programs will be readily available for the guidance and instructional functions of the school. As an important tool for keeping scores on permanent file and for presenting test results in a meaningful way, we have stressed use of some type of cumulative record card. Elementary-school and secondary-school records in frequent use in member schools of the Educational Records Bureau have been illustrated. Mention has been made of other types of test data, such as results on interest inventories, which should be filed in conjunction with the cumulative record cards.

Emphasis should be placed on keeping record cards up to date and maintaining the graphic portions of the record because of their particular value for assessing the overall picture of aptitude and attainment for the pupil. In some schools the personnel office will be able to keep the records up to date. In others, each teacher may be held responsible for recording the marks of objective test scores obtained by each pupil in his classes. In the interest of legibility and accuracy, the use of trained clerical help in maintaining the cumulative records is to be preferred. A special effort is usually required if the teachers and counselors are to make full use of the sections on personal characteristics, community activities, and pupil interests. These sections should not be neglected, for they are among the most important in the entire record.

SUGGESTIONS FOR FURTHER READING

1. Allen, Wendell C., *Cumulative Pupil Records*, New York, Bureau of Publications, Teachers College, Columbia University, 1943.
2. Ayer, Fred C., *Practical Child Accounting*, Austin, Texas, Steck Company, 1949.
3. Capehart, Bertis E., and Gore, Lillian L., "Cumulative Guidance Records," *School Executive*, 67:54-57 (November, 1948).

INTRODUCTION TO TESTING

4. Panzer, Joseph J., "The Cumulative Record—A Tool of Good Administration," *Catholic Educational Review*, 47:315-318 (May, 1949).
5. Traxler, Arthur E., *How to Use Cumulative Records: The Manual for the SRA Cumulative Record for Junior and Senior High Schools*, Chicago, Science Research Associates, 1947.
6. Traxler, Arthur E., *Techniques of Guidance*, New York, Harper & Brothers, 1945, chaps. 11, 12.
7. United States Office of Education, National Committee on Cumulative Records (David Segel, Chairman), *Handbook of Cumulative Records*, Federal Security Agency, Office of Education, Bulletin 1944, No. 5, Washington, United States Government Printing Office, 1945.

How Shall We Use Test Results?¹

TESTING in schools is usually undertaken in an effort to gain an understanding of the individual pupil and to adjust his educational program to his needs. Any use of test results growing out of this aim will probably be sound and defensible. Certain other uses may lead to misinterpretations unless precautions are taken. The uses of test results for guidance and the improvement of teaching are, in general, desirable. Difficulties sometimes arise when test results are used for certain administrative purposes.

Some of the more important uses of tests have already been discussed. In this chapter the main uses of tests will be summarized, and references will be made to various books and other publications which contain more extensive discussions.

PREREQUISITES TO EFFECTIVE USE OF TEST RESULTS

Before specific uses of test results are taken up, certain important prerequisites to effective use of the results of tests should be mentioned.

The first of these prerequisites is coöperation

¹ This chapter overlaps with some of the items taken up in preceding chapters, but it seemed desirable to include a chapter that would tie together ideas expressed in various places through the book.

between teachers and test specialists in test construction. Effective use of tests starts with the making of the tests. This statement does not mean that faculty members of local school systems should try to prepare tests comparable to standardized tests. There are now well-defined techniques of test construction, and seldom are school faculties sufficiently acquainted with these techniques to prepare reliable and valid tests without expert guidance. But coöperation of teachers in the test-making process is essential—particularly for achievement tests.

In the first place, teachers need to advise test makers on what important aspects of their own fields should be measured. For example, when the test-construction committees of the Educational Records Bureau in mathematics and science began their work, their first step was to send a questionnaire to teachers in Bureau member schools to find out what objectives and what course content these teachers believed to be important. Similarly, when the Committee on Diagnostic Reading Test, Inc., undertook the construction of tests for the analysis of reading difficulties, it first of all asked a large number of specialists in the field of reading what should be measured. The replies of teachers were extremely helpful in guiding the work of all these committees.

INTRODUCTION TO TESTING

In the second place, after preliminary drafts of test questions have been made, the criticism of teachers on each question is needed. Occasionally, items that have satisfactory statistical validities are not satisfactory from an instructional point of view. Even two or three questions in a test which do not meet with the approval of teachers may create an unfavorable impression concerning the entire test and thus impair its usefulness.

In the third place, teachers, as they use tests, need to provide test publishers and test service organizations with constructive criticisms and suggestions so that the experience of test users can be drawn upon when revised editions are published.

A second prerequisite to the effectiveness of test results is the planning and carrying on of systematic testing programs annually or semi-annually. One of the most important reasons why some schools do not use tests effectively is that they have no organized, systematic, coördinated testing program. The tests chosen for one year may have little relationship to those given the preceding year, and the tests given in one department of the school are interpreted by a different kind of norm from that used for tests administered in another department. This sort of haphazard testing is largely a waste of time for everyone concerned. The first basic rule in planning and carrying out a school-wide testing program is to make certain that the test results are comparable from year to year and from test to test. To be sure, new tests of proved value should be introduced into the testing program even though they are not precisely comparable to tests that have been used, but these innovations should be presented in such a way that they will cause the least possible interference with the comparability of the test results. For

example, the Committee on Tests and Measurements of the Educational Records Bureau annually makes specific recommendations to the member schools concerning tests to be used in the fall and spring programs. The committee makes a consistent effort to keep the tests comparable from year to year, from grade to grade, and from test to test. At the same time, new tests are carefully studied at each meeting of the committee, but a test which has been used is replaced by a new test only when the committee has evidence that it should do a better job.

A third prerequisite to the effective use of test results is inclusion of different types of tests in the testing program. A testing program based on just one type of test, such as an intelligence test or an achievement test battery, is of limited value. Preferably, the school-wide testing program should include (1) tests of general scholastic aptitude, (2) tests of achievement, and (3) inventories of interests. In the interpretation of test results, comparisons should be made among these three kinds of tests.

It may be questioned whether tests of specific aptitude and inventories of personality belong in a testing program for all pupils. Tests of specific aptitude, such as clerical ability tests or mechanical aptitude tests, are likely to be important in the guidance of some individuals but not in that of others. It would seem, therefore, that these tests should be administered on an individual, or small group, basis as needed. Inventories of personal qualities, likewise, should probably be reserved for supplementary testing on an individual basis. Their interpretation calls for a background of training in psychology, and the results are likely to be of doubtful validity if they are administered as part of a regular testing program. These instruments may be useful in an individual counseling situation where

HOW SHALL WE USE TEST RESULTS?

good rapport has been established and the interpretation of the results is in trained hands.

A fourth prerequisite to the valid use of test results is taking the tests "in stride." Teachers sometimes ask what they should do to prepare their pupils for the objective tests in a spring testing program. The answer is that they should do nothing at all, other than to carry on their regular class work and to make sure that any pupils who have not taken objective tests have an opportunity to see some sample questions of this kind. Above all, pupils should never be coached for the specific items in a test. If coaching takes place, the results will be worthless, and they may be actually harmful because they will tend to mislead all users of the results.

A fifth prerequisite to effective use of test results is accurate administration and scoring of tests. Not infrequently the use of test results is vitiated by careless administration of the tests and errors in scoring. The examiner must provide testing conditions under which pupils will be stimulated to do their best and must adhere strictly to the directions for administering the tests. The scoring needs to be done by competent, trained clerical workers, and provisions should be made for checking all operations. Scoring is such an exacting procedure that many schools prefer to use the services of an outside agency. Where machine scoring is needed, outside services may be almost a requirement. Even machine scoring, however, is likely to be inaccurate unless examiners use great care to see that the answer sheets are marked well with pencils containing electrographic lead.

A sixth prerequisite to the dependable use of test results is the organizing, recording, and reporting of test results in a form that is readily understandable and usable. Some schools that carry out the earlier stages of a testing program

well fail to organize the results so that they can be used by teachers and counselors. Scored test papers are of limited value in instruction and guidance. Before the results can be very useful, a staff of trained statistical and clerical workers must prepare from the test booklets distributions, class lists, profile charts, and cumulative records, must find class medians and quartiles (or means and standard deviations), and must present the results in terms of some kind of meaningful derived score, such as grade score or percentile. These procedures, like scoring, call for a great deal of detailed, highly accurate clerical work. As in the case of scoring, some schools prefer to utilize the services of organizations that specialize in this kind of work.

Samples of materials helpful in the interpretation and use of the test results have been shown in other chapters. These materials include especially distributions of scores, class lists, profile charts, cumulative records, and summary tables.

A seventh prerequisite to the use of test results is in-service training of teachers in interpretation and use of test results, including the minimum essentials of statistics. Each school needs someone on its staff who can assume responsibility for seeing that teachers and counselors have sufficient understanding of the elementary statistical concepts involved in testing to prevent their misunderstanding and misusing the results.

Only a very little in the way of statistics is necessary, but this small amount should be thoroughly learned. Every teacher should be acquainted with such fundamental concepts as the meaning of percentiles, the fact that percentiles in different parts of the scale are not equal, the fact that there is a probable error of measurement in every test score (even when

INTRODUCTION TO TESTING

the scoring is completely accurate) so that small differences in score are of no importance, and the fact that I.Q.'s or grade scores derived from different tests are not necessarily comparable. If gross misuse and misinterpretation of test results is to be avoided, there is no escape from the need to teach every teacher just a little statistics.

USES OF TEST RESULTS

There is a variety of reasons for the administration of tests in schools and colleges, and there are several important uses of test results. Of these, *the guidance use would seem to be the most important*. Historically, measurement and guidance have grown up together in the United States. Measurement without guidance loses much of its purpose; guidance without measurement loses its scientific character and becomes highly intuitive. The relation of measurement to guidance grows out of the simple thesis that in order to provide guidance services for an individual a counselor must first understand him and that objective appraisal is an essential element in that understanding.

For purposes of logical discussion, we sometimes separate guidance services into different areas, such as educational guidance, vocational guidance, and adjustment counseling. In practice, these areas are so closely related that when a counselor is working with an individual pupil it is often impossible for him to say which area he and the counselee are in. Not improbably they are working in all three areas at the same time. So a variety of tests is likely to have a bearing upon a particular guidance problem and to be of help in its solution.

Schools frequently do guidance testing—that is, testing for guidance purposes—but the term

“guidance test” would be a misnomer. The same tests that are used for other purposes—tests of scholastic aptitude, achievement, interests, specific aptitudes, and personal qualities—are used in guidance.

Some illustrations of ways in which test results are used in guidance are the following:

1. Study of aptitude and achievement test scores of a pupil in relation to norms for such groups as college freshmen, engineering sophomores, or employed accountants, in order to advise him concerning educational or vocational choice.
2. Study of relation of achievement test percentiles to scholastic aptitude percentiles in order to discover individuals who are not sufficiently motivated to do their best work or who are pushing themselves beyond their capacity. In this connection, the caution mentioned on page 66 should be observed.
3. Analysis of interest test profiles of individuals in order to identify areas of concentration of interests which may then be further investigated through the use of appropriate aptitude tests.
4. Use of high-school tests in predicting success on criteria, such as college-entrance tests, so that pupils may be counseled on the advisability of preparing for certain highly selective colleges. (For example, it has been found that the verbal score on the secondary Education Board Junior Scholastic Aptitude Test has an average correlation of approximately .80 with the verbal score on the College Entrance Examination Board Scholastic Aptitude Test, even after an interval of three or four years. Thus, the JSAT verbal score is a valuable predictor of success on SAT verbal score, even as early as Grade 8.)
5. Study of the all-round development of in-

HOW SHALL WE USE TEST RESULTS?

dividual pupils through the use of cumulative records, as illustrated in Chapter 8.

A second use of test results is for the purpose of individualization of instruction. Test results constantly remind schools of the large differences among individuals who have been assumed to be at the same level. Typically, on any test, such as reading comprehension or arithmetic reasoning, the differences between the medians for successive grades are small and the differences among the scores of the pupils at any grade level are many times as great. The test results thus point up to schools the necessity of trying to individualize instruction within the group and they provide an objective basis for starting differentiated instruction.

A third use of test results is in the diagnosis of the strengths and weaknesses of individual pupils and in either making allowance for or correcting weaknesses. One can, for example, administer the Yale Educational Aptitude Tests to an individual and obtain a reliable profile of scores in seven areas. It may be found that a certain pupil has low scores in the spatial visualizing and mechanical ingenuity areas, but very little can be done to correct these weaknesses. One can simply take account of them and try to guide the student into fields other than engineering and similar vocations where these aptitudes are needed.

Or, one may administer the Survey Section of the Diagnostic Reading Tests to a ninth-grade boy of better than average general intelligence and find that his reading comprehension is below the tenth percentile. This situation, if it is supported by other evidence, probably calls for remedial work.

The word "probably" in the preceding sentence is important. It should not be assumed that remedial work is needed until other factors,

such as health, emotional adjustment, and rate of growth in verbal intelligence, have been investigated. Any of these factors, or a combination of them may be basic to the reading difficulty. In particular, it is desirable to stress the importance of rate of mental growth, partly because this factor has been so neglected in the past. There is an accumulating body of evidence that mental growth curves of individuals are highly variable. What appears superficially to be a serious learning deficiency may be a temporary retardation in mental growth, which will probably straighten out later, and with less emotional tension, if the individual is not subjected to the pressure of remedial work. The point is that it is highly advisable to study the whole individual before deciding that he is a "remedial case."

A fourth use of test results is in the appraisal of the effectiveness of different kinds of instruction. The results of objective tests can sometimes be used in helping a school decide whether a certain kind of instruction is effective and whether changes should be introduced. This kind of use should be made only after there has been a careful study of the test in relation to the school's objectives and it should be cautiously applied in any event.

The point should be stressed that test results almost never should be used in evaluating individual teachers. In rare instances an accumulation of test results over a period of years or from different classes may suggest questions that should be looked into, but there are so many factors in any teaching situation that it is dangerous to try to draw definite conclusions on the basis of test results alone.

A fifth use of test results is in action research. Action research means research on the job to investigate questions of practical import as con-

INTRODUCTION TO TESTING

trusted with research into theoretical questions or eternal truths. This use of test results is closely related to the preceding one. In fact, action research is the only dependable way of investigating questions of instruction or curriculum by means of tests. For instance, the faculty of one public school wanted to know the value of an integrated curriculum in social studies at the tenth-grade level. The question was set up as an action research problem, carefully planned with regard to the experimental design. The results, which were favorable to the integrated curriculum, helped the school plan its future program in the social studies field. A much larger amount of this kind of research should be undertaken by schools.

A sixth use of test results is in counseling parents. This use might have been included under guidance, but since parents not infrequently present more difficult guidance problems than pupils do, it seems advisable to take special note of this use of tests. The parent may be inclined to disagree with the teacher's subjective mark or grade, particularly if the pupil is failing and, in talking with his parents, has perhaps blamed his failure on unfairness of the teacher. When attention is focused upon objective test results, however, the personal element is minimized. The parent can usually accept the fact that his boy ranks at, say, the fifth percentile in local norms on a machine-scored English test more easily than he can accept the grade of "F" given his boy by Miss Jones, the English teacher, on the basis of her opinion concerning his work.

Through discussion of the results of objective tests with parents, frequently better understanding of the capacities and limitations of the boy or girl will be brought about, and the school and the home may be able to work together toward developing the pupil's abilities. Parents are

known to be somewhat unrealistic occasionally in thinking of what the offspring should do or be. A conference in which a boy's cumulative test record is carefully explained to his parents and studied with them will often go a long way toward enlisting their coöperation in the choice of reasonable educational and vocational goals for the boy.

It is seldom advisable to send the results of tests home to parents in the way school grades are reported. Usually, it is best to bring in test results during conference so that careful explanation is possible. In situations where there is close coöperation between the school and parent groups in the community, and where a thorough program of informing parents concerning the meaning of test scores is carried on, it may be possible to educate the parent to receive and use wisely a direct report of results of aptitude and achievement tests, particularly if the program is fairly well standardized. As a general rule, neither intelligence quotients nor results of measures of personal qualities should be given to parents, either in written reports or in conferences, for there is especial danger that these kinds of data will be misunderstood.

A seventh use of test results is in reports to colleges and to prospective employers. Test results accumulated during several years of a student's schooling are to an increasing extent being given consideration by authorities concerned with the next higher educational level and by employers. Two illustrations may be cited in support of this point. Of 400 colleges whose replies to a questionnaire were summarized in the Fourth Report of the Committee on School and College Relations of the Educational Records Bureau, more than three-fourths stated that they would give full weight to comparable tests. In connection with an accounting

HOW SHALL WE USE TEST RESULTS?

testing program carried on by the Committee on Selection of Personnel of the American Institute of Accountants, more than 1300 employers—individuals and firms—in the field of public accounting said that they would give considerable weight to these tests in considering applicants for positions.

If a school carries on a regular, systematic testing program, it may not be necessary to give any tests specifically for reports to colleges or employers. The results of the same battery of tests that is administered for a variety of purposes within the school may be brought together when the individual reaches the end of his school course and used for objective reporting to the educational institution or employer receiving him after graduation.

In conclusion, it is desirable to stress the importance of considering the results of tests as one part of a pupil's total cumulative record,

both laterally and chronologically. Objective tests have occasionally been criticized on the ground that they cause a pupil to lose individuality and to become simply a point in a distribution. Nothing could be farther from the truth. When the results of a variety of tests over a period of years are brought together and recorded on a well-organized cumulative record, each little point (each score or percentile, which in itself has almost no meaning) takes on meaning from its place in and its relationship to the total pattern. As test results are added to the record, literally hundreds of interrelationships and combinations of the data in a single cumulative record become possible. As one studies these relationships, a living, growing individual emerges. Thus, test results, properly used, do not cause us to lose sight of individuals; rather, they help us to see these individuals more clearly, and as they really are.

SUGGESTIONS FOR FURTHER READING

1. Crawford, A. B., and Burnham, P. S., *Forecasting College Achievement*, New Haven, Yale University Press, 1946, pp. 134-169.
2. Darley, John G., *Testing and Counseling in the High School Guidance Program*, Chicago, Science Research Associates, 1943, pp. 24-41.
3. Froehlich, Clifford P., and Benson, Arthur L., *Guidance Testing*, Chicago, Science Research Associates, 1948, pp. 53-79.
4. Greene, Harry A., Jorgensen, Albert N., and Gerberich, J. Raymond, *Measurement and Evaluation in the Secondary School*, New York, Longmans, Green and Company, 1944, pp. 608-622.
5. Lindquist, E. F. (ed.), *Educational Measurement*, Washington, American Council on Education, 1950, pp. 47-67.
6. Traxler, Arthur E., *Techniques of Guidance*, New York, Harper & Brothers, 1945, pp. 185-201.
7. Traxler, Arthur E., *The Use of Test Results in Diagnosis and Instruction in the Tool Subjects*, Educational Records Bulletin No. 18, New York, Educational Records Bureau (rev. ed.), 1949.
8. Wood, Ben D., and Haefner, Ralph, *Measuring and Guiding Individual Growth*, New York, Silver Burdett Company, 1948.

How Does All This Apply to a Specific Case?

EDWARD MANSFIELD is one of thirty-three pupils entering Grade 7 of the Newtonville Junior High School in the fall of 1946. His transfer record from an out-of-state school indicates that he is an average-to-superior student and that his grades have been satisfactory. This is about all the information which his new teachers have concerning him. He appears rather tall for his age and somewhat thin for his height. He seems a little shy, but he is friendly and the boys in the seventh grade seem to like him.

During the first days of the school year objective tests are given to all pupils in the Newtonville schools. The Otis Self-Administering Test of Mental Ability, Intermediate Examination, Form B, and the Traxler Silent Reading Test, Form 1, are administered to the pupils in Grade 7.

When the test results are reported, it is found that Edward has obtained a raw score of 50 on the Otis test, equivalent to a mental age of 13-10. Since Edward's birth date is September 18, 1934, his chronological age was 12 years at the time the test was given. By means of the procedure outlined in the Otis manual for relating

the mental age to the chronological age, an I.Q. of 111 is derived.

No record of previous testing is available for Edward, so, in order to make sure of his ability level, the home-room teacher refers him to his counselor for an individual intelligence test. The Terman-Merrill Revision of the Stanford-Binet Scale is administered, and the results place Edward's I.Q. at 114. Since above-average aptitude has been shown on both tests, the teacher now feels secure in concluding that Edward is capable of doing satisfactory work. She knows that many factors other than academic aptitude may affect achievement, but if Edward's work is below average the teacher will feel fairly sure that the explanation lies in some area other than that of ability.

The test report shows the following results for Edward on the Traxler Silent Reading Test:

Traxler Silent Reading Test, Form 1			
Part	Score	P.S. %ile	P.S. Grade Equivalent
Reading Rate	30	57	
Total Comprehension	46	61	
Total Reading	76	58	7.9

HOW DOES ALL THIS APPLY TO A SPECIFIC CASE?

In observing this record, the teacher is interested to note that Edward's reading skills as measured by this test are slightly above average when compared with those of other public-school pupils at the same grade level and that he ranks about the same in rate and comprehension. The grade rating for total score suggests that he is several months ahead of the average pupil entering Grade 7, but the degree of advancement is not noteworthy. These results are generally in line with the boy's aptitude results—perhaps just a little lower than one might expect on the basis of obtained I.Q.'s.

No other objective tests are given to Edward until the following spring. His teachers have in the meantime found him to be interested in dramatics, and his midyear grades indicate that he has been doing good work in English and spelling, that he does well in physical education, and that he has passing marks in his other subjects. As a result of two conferences with his class counselor, information concerning his home and parents, as well as other personal data, have been obtained for use by his teachers. During the winter Edward was absent for two weeks for a tonsillectomy and following this he had several severe colds. His teachers have observed that he has needed prodding rather frequently, that his influence in the class group is neutral, and that he is slow in adjusting to the new school situation.

The results of the Stanford Achievement Test which Edward took in the spring are shown on the profile in Figure 20. His test scores seem to agree rather well with his school marks. His best scores are in the parts of the test dealing with word meaning, paragraph meaning, and spelling, while his poorest performance is in social studies, arithmetic computation, and elementary science. On the whole, his general achievement

tends to run rather close to the average, somewhat lower than might be expected on the basis of his aptitude test results. He has a total grade rating of 8.1 on the test.

The picture which the record of school work and test information obtained thus far makes when it is assembled as the beginning of a cumulative record is illustrated in Figure 21. Note that the Otis score and mental age are entered in the tabular portion of the card which presents the results of various group tests, but that the Otis I.Q. is shown on the reverse side (Fig. 22) along with the I.Q. from the Stanford-Binet Scale. The I.Q.'s are recorded on the other side of the card so that they need not be revealed when the achievement test portion of the record is being discussed with the pupil, his parents, or others who might misinterpret intelligence quotients.

The school marks shown above the test record indicate passing work in all subjects, but his home-room teacher has commented that there is need for more conscientious work in social studies. His work in arithmetic improved from January to June.

The information entered on the other side of the cumulative record as a result of counseling interviews is shown in Figure 22. It is apparent from these entries that Edward's first year in the Newtonville Junior High School has not been characterized by completely satisfactory adjustment to the school program. The counselor has noted that home cooperation is needed in bringing about better school adjustment.

The next fall, as Edward begins the work of the eighth grade, he is given academic aptitude and reading tests once more. Different forms of the Otis and Traxler tests are employed. Very little change in either academic aptitude or percentile ranking of reading skills is observed

EDUCATIONAL PROFILE CHART - STANFORD ACHIEVEMENT TEST

ADV

FORM D

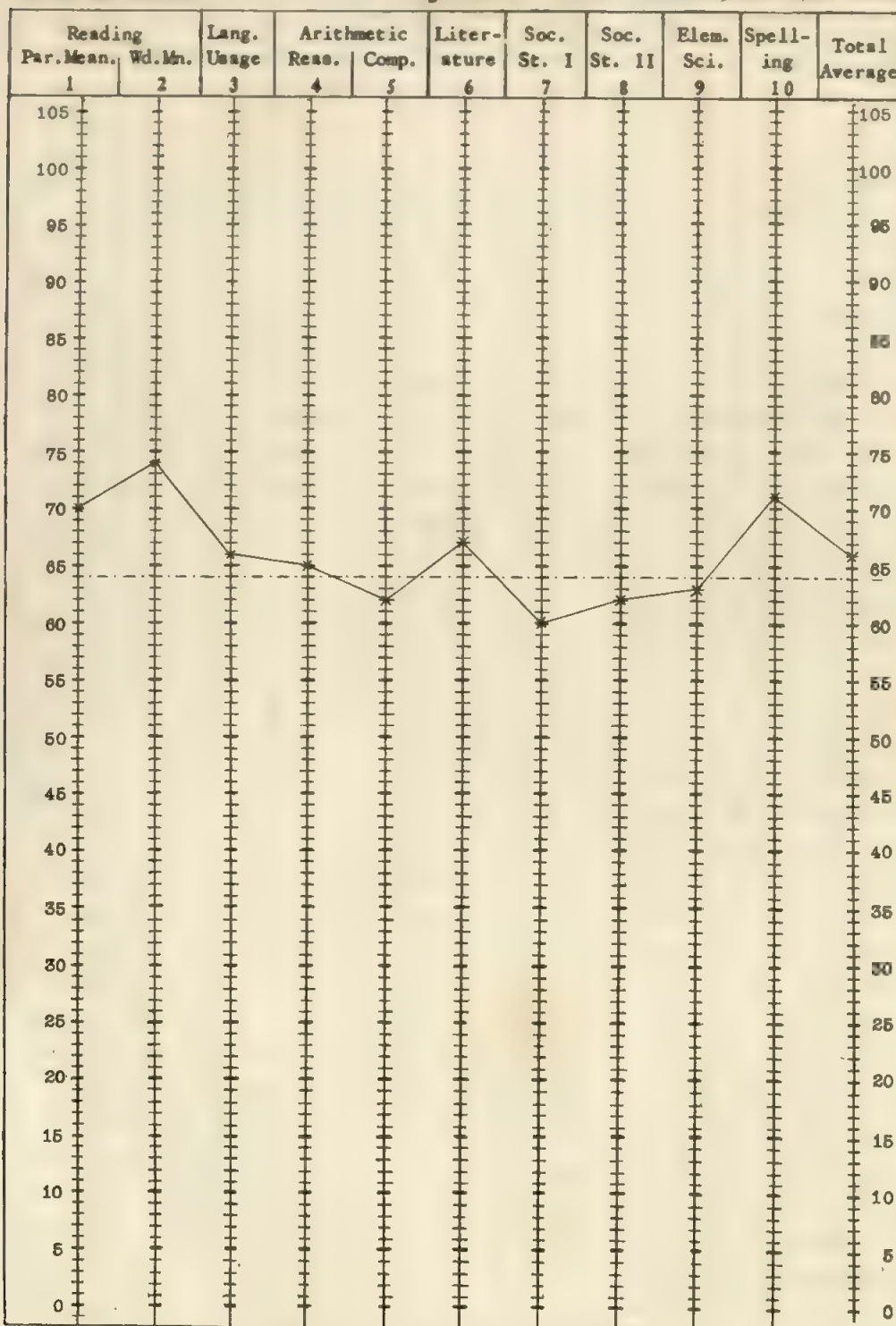
Name MANSFIELD, EDWARD S.

School NEWTONVILLE JUNIOR HIGH SCHOOL

Grade 7

Age 12-7

Date APRIL, 1947



Test	Score	P. S. Gr.
1		
2		
Av. Rd.		
3		
4		
5		
Av. Ar.		
6		
7		
8		
Av. S. S.		
9		
10		
Total		
P. A.		

Age Grade *

15-0	11-0
14-0	10-0
13-0	9-0
12-0	8-0
11-0	7-0
10-0	6-0
9-0	5-0
8-0	4-0
7-0	3-0
6-0	2-0
5-0	1-0
4-0	0-0
3-0	0-0
2-0	0-0
1-0	0-0
0-0	0-0

Gr. 7

* Grade defined as in Table 2 of the *Directions for Administering*.

** Educational ages above 15-0 and below 7-9 are extrapolated.

FIGURE 20. Stanford Achievement Test Profile of Edward S. Mansfield. (This profile form is adapted by the Educational Records Bureau from the published form for this test with special permission of World Book Company.)

HOW DOES ALL THIS APPLY TO A SPECIFIC CASE?

when the new scores are compared with the seventh-grade record. The Otis I.Q. obtained in the fall of 1947 is 108 as compared with one of 111 for the preceding fall, and the total reading percentile has risen from 58 to 65.

At the end of the first semester Edward's academic record is about the same as in Grade 7. He seems to be having considerable difficulty with science but maintains passing grades in social studies and shows above-average work in English, arithmetic, and physical education. His performance in music is satisfactory without being outstanding.

Notes added to the cumulative record by the counselor indicate that Edward has taken on a newspaper delivery route and that coöperation has been obtained from the home with regard to the boy's school adjustment.

In the spring, the Stanford Achievement Test is administered to all pupils in Grade 8, and again Edward's results indicate about the same pattern of achievement as that displayed near the end of Grade 7. In order to study the amount and direction of growth more accurately, Edward's home-room teacher plotted his Stanford results on an educational profile chart together with his seventh-grade profile. The two profiles are shown in Figure 23.

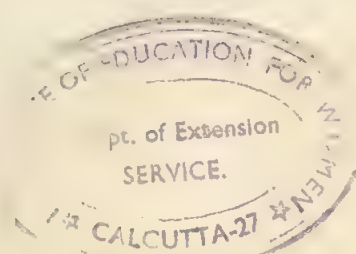
When the profiles for two years are compared, it is apparent that Edward has shown more growth in some areas than in others. He shows less advancement in reading than in some of the other subjects, although in this section of his profile he continues to be higher than in most of the other recorded scores. His greatest growth is in social studies and in arithmetic computation. By checking the grade equivalents at the right of the profile which correspond to the various scores one can determine that his growth in these areas represents approximately a year and

a half of advancement in terms of the public-school grade norms. His least growth is in elementary science. The teacher views this as being particularly unsatisfactory, since formal instruction in elementary science was started at the beginning of Grade 8, and she feels that this instruction should yield considerable increase in knowledge of this subject matter. His advancement in general achievement as reflected in total average score is just about one year of grade growth.

At the end of the school year 1947-48, the cumulative record entries appear as in Figures 24 and 25. Study of this record indicates that Edward has continued his work with the Boy Scouts and with the dramatics club, that his health is considerably improved, and that there is evidence of growth in social adjustment and acceptance of responsibility. His work in the dramatics club has received recognition. His academic record is about the same as in Grade 7, with some improvement shown in social studies. At first he had considerable difficulty in science, but his work here was improved with intensive study.

As the cumulative record was extended through Grades 9 and 10 it became apparent that Edward was making a much better adjustment to school. Counselors' notes reveal steady improvement in health, personal adjustment, and social adjustment. His interests in debating and physical education were noticeable, and his liking for dramatics continued.

When Edward entered the Newtonville Senior High School in the fall of 1949 his cumulative record was transferred to the senior-high-school principal's office. He made the transition from the junior to the senior high school without difficulty. Tests administered in Grade 10 included a test of primary mental abilities. The



NAME MANSFIELD, EDWARD SAMUEL

BIRTH DATE SEPTEMBER 18, 1934

SEX: (M) F

RACE: W N Y I

YEAR	SCHOOL	GRADE	CHRON. AGE	NOTES ON ELEMENTARY SCHOOL EXPERIENCE.																				
1946-1947	NEWTONVILLE Junior High	7	12																					
FIELD OF STUDY					FIELD OF STUDY					FIELD OF STUDY					FIELD OF STUDY					FIELD OF STUDY				
ENGLISH	B	C	I		TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile
SPELLING	B	A	I																					
ARITHMETIC	C	B	I																					
GEOGRAPHY	C	C	I																					
HISTORY	C	C	I																					
MUSIC	S	S	1/2																					
DAYS EQU.	B	B	1/2																					
NEED FOR MORE CONSCIENTIOUS WORK IN GEOGRAPHY AND HISTORY.																								
RECORD OF CLASS WORK					COMMENT																			
TEST	Sc.	M.A.	% ile*		TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile
OCTOBER - TEST	76	79	83																					
NORVILL STATE	30	30	37																					
TEST	Sc.	P.S. Gr.	% ile*		TEST	Sc.	P.S. Gr.	% ile	TEST	Sc.	P.S. Gr.	% ile	TEST	Sc.	P.S. Gr.	% ile	TEST	Sc.	P.S. Gr.	% ile	TEST	Sc.	P.S. Gr.	% ile
STAN BRUNDA	66	81																						
ONE MEIN.	70	90																						
WORD MEIN.	74	100																						
LONG VETER	66	81																						
BIRTH DEVS	65	78																						
ALVIN CAMP	62	72																						
LITERATURE	67	83																						
SEC. ST. I	60	68																						
SEC. ST. II	62	72																						
FLEET SEC.	63	74																						
SPELLING	71	93																						
ACHIEVEMENT TESTS AND OTHER TESTS																								

* Indicate basic Group (public school, independent school, local school, etc.)

NOTES

PRINCIPAL'S RECOMMENDATION:

This space reserved for notes by College Admissions Officer or Employer.

FIGURE 21. Cumulative Record of Edward S. Mansfield in Grade 7 (Front of Card). (This is a simplified SRA Cumulative Record Form prepared by Arthur E. Traxler, published by Science Research Association, 1947. Used with permission of Science Research Associates.)

NAME <i>MANSFIELD, EDWARD JAMES</i>		BIRTH DATE <i>9-18-34</i>	BIRTHPLACE <i>JOHNSONVILLE, RELIGION METHODIST</i>	ADDRESS <i>28 LOCKOUT AVE. NEWTOWN</i>	TELEPHONE <i>6-1428</i>
FAMILY	FATHER <i>Howard A.</i>	OCCUPATION	EDUCATION	RELIGION	HEALTH
	MOTHER <i>Audrey S.</i>	<i>REAL ESTATE DEALER</i>	<i>2 YEARS COLLEGE BROWN UNIVERSITY</i>	<i>METHODIST</i>	<i>Good</i>
	STEP PARENT OR GUARDIAN	<i>HOUSEWIFE</i>		<i>METHODIST</i>	<i>Good</i>
	SIBLINGS				
	<i>William Jarvis</i>	<i>ACCOUNTANT</i>	<i>B.A. ACCOUNTING UNIV. OF MIAHI</i>	<i>BIRTH DATE</i>	
	<i>MARY RUTH</i>	<i>STUDENT</i>	<i>NOW ATTENDING HUNTER COLLEGE</i>	<i>4-23-24</i>	<i>EXCELLENT</i>
			<i>11-22-30</i>	<i>Good</i>	
YEAR AND AGE	<i>1946-1947</i>	<i>C.A. 12</i>			
INTELLIGENCE QUOTIENT	<i>Binet III, Otis III</i>				
COUNSELOR	<i>MISS WARRIN</i>				
ATTENDANCE	<i>ABSENT TWO WEEKS FOR TONSILLECTOMY</i>				
ACADEMIC DISCIPLINE	<i>A LITTLE SLOW IN ADJUSTING TO JR. H.S. PROGRAM - NEW SCHOOL.</i>				
PERSONAL					
IN SCHOOL	<i>DRAMATICS CLUB</i>				
INTERESTS					
OUT OF SCHOOL	<i>BOY SCOUTS</i>				
VOCATIONAL EXPERIENCES	<i>NONE</i>				
NOTABLE ACCOMPLISHMENTS	<i>DID WELL IN PART IN PLAY GIVEN BY DRAMATICS CLUB IN ASSEMBLY.</i>				
HEALTH AND PHYSICAL VIGOR	<i>QUITE SUSCEPTIBLE TO COLDS. FOLLOWING TONSILLECTOMY - RECOVERY RATHER SLOW.</i>				
EDUCATIONAL AND VOCATIONAL PLANS	<i>NO PLANS AS YET.</i>				
RESPONSIBILITY	<i>HAS NEEDED PROMOTING RATHER FREQUENTLY</i>				
DEPENDABILITY					
CREATIVENESS AND IMAGINATION	<i>SOME ORIGINALITY, BUT SEEMS TO NEED DIRECTION</i>				
INFLUENCE	<i>NEUTRAL</i>				
SOCIAL ADJUSTMENT	<i>HAS ADJUSTED TO NEW GROUP SLOWLY.</i>				
COUNSELOR'S NOTES	<i>NEED TO VISIT HOME TO SECURE COOPERATION TOWARD BETTER SCHOOL ADJUSTMENT.</i>				
REMARKS					

Note here anything of significance concerning the family or conditions in the home.

BEHAVIOR DESCRIPTION

FIGURE 22. Cumulative Record of Edward S. Mansfield in Grade 7 (Back of Card).

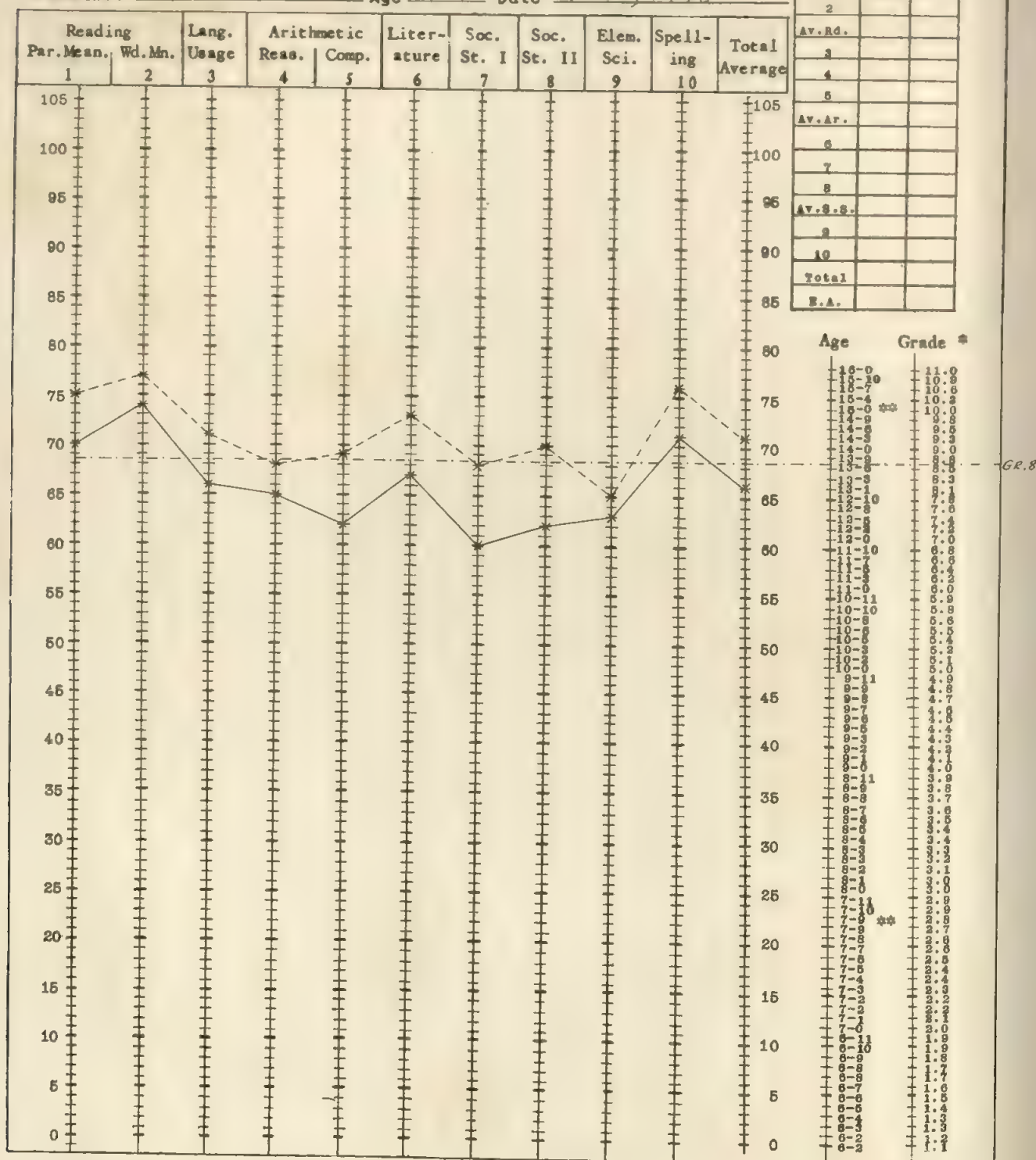
Name MANSFIELD, EDWARD S.School NEWTONVILLE JUNIOR HIGH SCHOOLGrade 8Age 13-7 Date APRIL, 1948

FIGURE 23. Stanford Achievement Test Profile of Edward S. Mansfield in Grade 8 Compared with His Profile in Grade 7. (This profile form is adapted by the Educational Records Bureau from the published form for this test with special permission of World Book Company.)

HOW DOES ALL THIS APPLY TO A SPECIFIC CASE?

results indicated outstanding aptitude in verbal meaning with comparatively high aptitude in word fluency and reasoning but below-average scores for number and space. Scores on the American Council on Education Psychological Examination substantiate this trend. On the linguistic section of the latter test, Edward surpassed 82 percent of the pupils of corresponding grade level, while in quantitative aptitude or Q-score he exceeded only 32 percent of the norm group.

In January, 1951, Edward's brother died of injuries received in an automobile accident. Considerable upset in the family resulted, partly because the brother remained alive but in a critical condition for some two weeks following the accident and partly because the mother was grief-stricken for a long period over the loss of her eldest son. The family situation had some noticeable effects on Edward's school work during the spring of 1951. The mother made a slow readjustment, and by the end of the school year Edward was again well adjusted to his group and was doing school work generally at a satisfactory level.

Edward's test record continued to show good ability in verbal areas, and his record of academic progress bore out these test scores. His grades were particularly good in English and French and in verbal subjects such as economics, commercial law, and business practice. His academic record continued satisfactory in all subjects. He developed into a good athlete and earned varsity letters in two sports. His grades in physical education reflected his athletic prowess. The counselor's notes indicated that responsibility and leadership were becoming noticeable in Edward's school activities and that he was well liked by his group. Outstanding work in debating and dramatics and excellent

achievement in economics, business, and commercial law directed Edward's interests toward the profession of law. His vocational experience as a clerk in a law office during the summer of 1951 encouraged this interest. Near the end of the senior year he had developed a definite plan to enter the state university and prepare for the law profession.

At the end of the senior year, the cumulative record for this boy appears as shown in Figures 26 and 27. The case summary made by Edward's counselor in the spring of 1952 and sent with a copy of the cumulative record card to the state university is as follows:

Edward Mansfield was graduated in the upper half of his class in the spring of 1952 after attending the Newtonville Junior-Senior High School for a period of six years. He came to us at the beginning of Grade VII from the public elementary schools in Johnsonville, Florida. His preparation in the lower grades was sufficient to enable him to go on with our regular seventh grade group, although he showed some weaknesses in social studies and arithmetic which required some time to correct. His social adjustment and physical development have been satisfactory throughout the period of his schooling at Newtonville. Edward's academic aptitude, as measured by several tests, is somewhat above average. He has particular skill in linguistic and verbal areas and somewhat less ability in the field of mathematics. His interests are in the areas of business and law.

This pupil's achievement, as shown by his school marks and confirmed by scores on objective tests, is somewhat above average. His best efforts have been in English, economics, law, business, and physical education. His success in those parts of the curriculum more verbal in nature is in line with his interests. His occupational choice at present is the profession of law.

The Mansfield family is an intelligent, cooperative one. His mother has had some difficulty in connection with a death in the family but recovery has

NAME MANSFIELD, Edward Samuel
BIRTH DATE 9-18-34
BIRTHPLACE Tomsboro, NJ, Religion Methodist
ADDRESS 28 Woodmont Ave., Newton, NJ 07428
TELEPHONE 6-1428
1430 W. Forest, Newton, NJ 07428

FAMILY	OCCUPATION	EDUCATION	RELIGION	HEALTH	DECEASED DATE
FATHER HOWARD A.	REAL ESTATE DEALER	2 YEARS COLLEGE BROWN UNIVERSITY	METHODIST	GOOD	
MOTHER AUDREY S.	HOUSEWIFE		METHODIST	GOOD	
STEP PARENT OR GUARDIAN					
SIBLINGS					
WILLIAM JARVIS	ACCOUNTANT	B.A. ACCOUNTING UNIV. OF MICH.	BIRTH DATE	EXCELLENT	
MARY RUTH	STUDENT	NOW ATTENDING HUNTER COLLEGE	4-23-24 11-22-30	GOOD	
YEAR AND AGE	1946-1947	C.A. 12	1947-1948	C.A. 13	
INTELLIGENCE QUOTIENT	BINET III, OTIS III		OTIS 108		
COUNSELOR	MISS WARRIN		MR. SNIDER		
ATTENDANCE	ABSENT TWO WEEKS FOR TONSILLITIS		REGULAR		
ACADEMIC DISCIPLINE	A LITTLE SLOW IN ADJUSTING TO J.R.N.S. PROGRAM - NEW SCHOOL		GOOD		
PERSONAL IN SCHOOL	DRAMATICS CLUB		DRAMATICS CLUB		
INTERESTS OUT OF SCHOOL	BOY SCOUTS		BOY SCOUTS		
VOCATIONAL EXPERIENCES	NONE		NEWSPAPER DELIVERY ROUTE		
NOTABLE ACCOMPLISHMENTS	DID WELL IN PART IN PLAY GIVEN BY DRAMATICS CLUB IN ASSEMBLY		VERY GOOD WORK IN DRAMATICS CLUB		
HEALTH AND PHYSICAL VIGOR	QUITE SUSCEPTIBLE TO COLDS FOLLOWING CONSILITIS - RECOVERY RATHER SLOW		HEALTH CONSIDERABLY IMPROVED OVER LAST YEAR. RATHER ACTIVE BOY		
EDUCATIONAL AND VOCATIONAL PLANS	NO PLANS AS YET		NO PLANS BEYOND HIGH SCHOOL		
RESPONSIBILITY DEPENDABILITY	HAS OFFERED READING RATHER FREQUENTLY		HAS ACCEPTED RESPONSIBILITY FOR WORK MORE DEMANDABLE THAN LAST YEAR		
CREATIVENESS AND IMAGINATION	SOME ORIGINALITY; BUT SEEMS TO NEED DIRECTION		HAS SHOWN ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		
INFLUENCE	NEUTRAL		GENERALLY CONSTRUCTIVE BEHAVIOR		
SOCIAL ADJUSTMENT	HAS ADJUSTED TO NEW GROUP SMOOTHLY		GETS ALONG WITH CLASS - NOTES FIRSTLY WELL LESS RETIRED THAN LAST YEAR		
COUNSELOR'S NOTES	NEED TO VISIT HOME TO SECURE COOPERATION TOWARD BETTER SCHOOL ADJUSTMENT		VISITED HOME PARENTS QUITE COOPERATIVE. EDWARD IS SHOWING MORE INTEREST ADJUSTMENT TO SCHOOL		
REMARKS					

Figure 25. Cumulative Record of Edward S. Mansfield in Grades 7 and 8 (Back of Card).

NAME MANSFIELD, EDWARD SAMUEL

BIRTH DATE SEPTEMBER 18, 1934

SEX: M F

RACE: W N Y I

YEAR	1946 - 1947	1947 - 1948	1948 - 1949	1949 - 1950	1950 - 1951	1951 - 1952																																																																																																																																																																																																																																																
SCHOOL	NEWTONVILLE JUNIOR HIGH																																																																																																																																																																																																																																																					
GRADE	7																																																																																																																																																																																																																																																					
CHRON. AGE	12																																																																																																																																																																																																																																																					
NOTES ON ELEMENTARY SCHOOL EXPERIENCE.	<table border="1"> <thead> <tr> <th>FIELD OF STUDY</th> <th>JAN</th> <th>JUNE</th> <th>CR</th> <th>FIELD OF STUDY</th> <th>JAN</th> <th>JUNE</th> <th>CR</th> <th>FIELD OF STUDY</th> <th>JAN</th> <th>JUNE</th> <th>CR</th> <th>FIELD OF STUDY</th> <th>JAN</th> <th>JUNE</th> <th>CR</th> <th>FIELD OF STUDY</th> <th>JAN</th> <th>JUNE</th> <th>CR</th> </tr> </thead> <tbody> <tr> <td>ENGLISH</td> <td>B</td> <td>B</td> <td>1</td> <td>ENGLISH</td> <td>B</td> <td>B</td> <td>1</td> <td>ENGLISH</td> <td>A</td> <td>B</td> <td>1</td> <td>ENGLISH</td> <td>A</td> <td>B</td> <td>1</td> <td>ENGLISH</td> <td>A</td> <td>B</td> <td>1</td> </tr> <tr> <td>SPELLING</td> <td>B</td> <td>A</td> <td>1</td> <td>ORTHOGRAHY</td> <td>B</td> <td>B</td> <td>1</td> <td>ELEN HIG.</td> <td>B</td> <td>B</td> <td>1</td> <td>FRENCH</td> <td>B</td> <td>B</td> <td>1</td> <td>GREEN (GVS)</td> <td>A</td> <td>A</td> <td>1</td> </tr> <tr> <td>ARITHMETIC</td> <td>C</td> <td>B</td> <td>1</td> <td>ELEM. SCI.</td> <td>D</td> <td>C</td> <td>1</td> <td>GEN. SCI.</td> <td>C</td> <td>C</td> <td>1</td> <td>INTER. ALC</td> <td>C</td> <td>C</td> <td>1</td> <td>PHYSICS</td> <td>C</td> <td>C</td> <td>1</td> </tr> <tr> <td>GEOGRAPHY</td> <td>C</td> <td>C</td> <td>1</td> <td>GEOGRAPHY</td> <td>C</td> <td>C</td> <td>1</td> <td>SOC. STUDIES</td> <td>B</td> <td>C</td> <td>1</td> <td>FUR. HIST</td> <td>B</td> <td>C</td> <td>1</td> <td>AM. HIST</td> <td>B</td> <td>B</td> <td>1</td> </tr> <tr> <td>HISTORY</td> <td>C</td> <td>C</td> <td>1</td> <td>HISTORY</td> <td>C</td> <td>B</td> <td>1</td> <td>ART</td> <td>A</td> <td>A</td> <td>1</td> <td>ECONOMICS</td> <td>B</td> <td>B</td> <td>1</td> <td>AM. HIST</td> <td>A</td> <td>A</td> <td>1</td> </tr> <tr> <td>MUSIC</td> <td>S</td> <td>S</td> <td>1</td> <td>MUSIC</td> <td>S</td> <td>S</td> <td>1</td> <td>MUSIC</td> <td>S</td> <td>S</td> <td>1</td> <td>PHYS. EDUC.</td> <td>A</td> <td>A</td> <td>1</td> <td>PHYS. EDUC.</td> <td>A</td> <td>A</td> <td>1</td> </tr> <tr> <td>PHYS. EDUC.</td> <td>B</td> <td>B</td> <td>1</td> <td>PHYS. EDUC.</td> <td>B</td> <td>B</td> <td>1</td> <td>PHYS. EDUC.</td> <td>B</td> <td>B</td> <td>1</td> <td>PHYS. EDUC.</td> <td>A</td> <td>A</td> <td>1</td> <td>PHYS. EDUC.</td> <td>A</td> <td>A</td> <td>1</td> </tr> </tbody> </table>						FIELD OF STUDY	JAN	JUNE	CR	FIELD OF STUDY	JAN	JUNE	CR	FIELD OF STUDY	JAN	JUNE	CR	FIELD OF STUDY	JAN	JUNE	CR	FIELD OF STUDY	JAN	JUNE	CR	ENGLISH	B	B	1	ENGLISH	B	B	1	ENGLISH	A	B	1	ENGLISH	A	B	1	ENGLISH	A	B	1	SPELLING	B	A	1	ORTHOGRAHY	B	B	1	ELEN HIG.	B	B	1	FRENCH	B	B	1	GREEN (GVS)	A	A	1	ARITHMETIC	C	B	1	ELEM. SCI.	D	C	1	GEN. SCI.	C	C	1	INTER. ALC	C	C	1	PHYSICS	C	C	1	GEOGRAPHY	C	C	1	GEOGRAPHY	C	C	1	SOC. STUDIES	B	C	1	FUR. HIST	B	C	1	AM. HIST	B	B	1	HISTORY	C	C	1	HISTORY	C	B	1	ART	A	A	1	ECONOMICS	B	B	1	AM. HIST	A	A	1	MUSIC	S	S	1	MUSIC	S	S	1	MUSIC	S	S	1	PHYS. EDUC.	A	A	1	PHYS. EDUC.	A	A	1	PHYS. EDUC.	B	B	1	PHYS. EDUC.	B	B	1	PHYS. EDUC.	B	B	1	PHYS. EDUC.	A	A	1	PHYS. EDUC.	A	A	1																																																																																
FIELD OF STUDY	JAN	JUNE	CR	FIELD OF STUDY	JAN	JUNE	CR	FIELD OF STUDY	JAN	JUNE	CR	FIELD OF STUDY	JAN	JUNE	CR	FIELD OF STUDY	JAN	JUNE	CR																																																																																																																																																																																																																																			
ENGLISH	B	B	1	ENGLISH	B	B	1	ENGLISH	A	B	1	ENGLISH	A	B	1	ENGLISH	A	B	1																																																																																																																																																																																																																																			
SPELLING	B	A	1	ORTHOGRAHY	B	B	1	ELEN HIG.	B	B	1	FRENCH	B	B	1	GREEN (GVS)	A	A	1																																																																																																																																																																																																																																			
ARITHMETIC	C	B	1	ELEM. SCI.	D	C	1	GEN. SCI.	C	C	1	INTER. ALC	C	C	1	PHYSICS	C	C	1																																																																																																																																																																																																																																			
GEOGRAPHY	C	C	1	GEOGRAPHY	C	C	1	SOC. STUDIES	B	C	1	FUR. HIST	B	C	1	AM. HIST	B	B	1																																																																																																																																																																																																																																			
HISTORY	C	C	1	HISTORY	C	B	1	ART	A	A	1	ECONOMICS	B	B	1	AM. HIST	A	A	1																																																																																																																																																																																																																																			
MUSIC	S	S	1	MUSIC	S	S	1	MUSIC	S	S	1	PHYS. EDUC.	A	A	1	PHYS. EDUC.	A	A	1																																																																																																																																																																																																																																			
PHYS. EDUC.	B	B	1	PHYS. EDUC.	B	B	1	PHYS. EDUC.	B	B	1	PHYS. EDUC.	A	A	1	PHYS. EDUC.	A	A	1																																																																																																																																																																																																																																			
COMMENT	<p>NEED FOR MORE CONSCIENTIOUS WORK IN GEOGRAPHY AND HISTORY.</p> <p>CONSIDERABLE DIFFICULTY IN SCIENCE; SOMEWHAT IMPROVED WITH INTENSIVE STUDY.</p> <p>IMPROVED WORK IN SCIENCE BUT STILL NEEDS WORK IN ART.</p> <p>GOOD WORK IN FRENCH. ECONOMICS AND CON-MERCIAL LAW CONTINUED DIFFICULTY WITH MATH. RECEIVED LETTERS IN FOOTBALL AND QUARTERBACK.</p> <p>VERY GOOD WORK IN ECONOMICS AND CON-MERCIAL LAW CONTINUED DIFFICULTY WITH MATH. RECEIVED LETTERS IN FOOTBALL AND QUARTERBACK.</p> <p>GOOD. MEDICALS IN PHYSICS 1/3 AN OUTSTANDING ACHIEVEMENT.</p>																																																																																																																																																																																																																																																					
ACADEMIC APTITUDE	<table border="1"> <thead> <tr> <th>TEST</th> <th>Sc.</th> <th>M.A.</th> <th>% ile</th> <th>TEST</th> <th>Sc.</th> <th>M.A.</th> <th>% ile</th> <th>TEST</th> <th>Sc.</th> <th>M.A.</th> <th>% ile</th> <th>TEST</th> <th>Sc.</th> <th>M.A.</th> <th>% ile</th> </tr> </thead> <tbody> <tr> <td>OTIS S.A. TEST</td> <td>50</td> <td>13.0</td> <td>11-4</td> <td>OTIS S.A. TEST</td> <td>36</td> <td>11-4</td> <td>15-10</td> <td>ALC. PAPER</td> <td>76</td> <td>70</td> <td>63</td> <td>ALC. PAPER</td> <td>76</td> <td>70</td> <td>63</td> </tr> <tr> <td>TEST</td> <td>Sc.</td> <td>P.S.</td> <td>% ile</td> <td>TEST</td> <td>Sc.</td> <td>P.S.</td> <td>% ile</td> <td>TEST</td> <td>Sc.</td> <td>P.S.</td> <td>% ile</td> <td>TEST</td> <td>Sc.</td> <td>P.S.</td> <td>% ile</td> </tr> <tr> <td>TRAILER - TOT</td> <td>76</td> <td>79</td> <td>58</td> <td>TRAILER - TOT</td> <td>91</td> <td>96</td> <td>65</td> <td>TRAILER - TOT</td> <td>91</td> <td>96</td> <td>65</td> <td>TRAILER - TOT</td> <td>91</td> <td>96</td> <td>65</td> </tr> <tr> <td>FORM 1 - COMP</td> <td>30</td> <td>30</td> <td>57</td> <td>FORM 1 - COMP</td> <td>40</td> <td>40</td> <td>67</td> <td>FORM 1 - COMP</td> <td>40</td> <td>40</td> <td>67</td> <td>FORM 1 - COMP</td> <td>40</td> <td>40</td> <td>67</td> </tr> <tr> <td>TEST</td> <td>Sc.</td> <td>P.S.</td> <td>% ile</td> <td>TEST</td> <td>Sc.</td> <td>P.S.</td> <td>% ile</td> <td>TEST</td> <td>Sc.</td> <td>P.S.</td> <td>% ile</td> <td>TEST</td> <td>Sc.</td> <td>P.S.</td> <td>% ile</td> </tr> <tr> <td>STAN. HEN. AND</td> <td>66</td> <td>81</td> <td>81</td> <td>STAN. HEN. AND</td> <td>71</td> <td>93</td> <td>93</td> <td>STAN. HEN. AND</td> <td>71</td> <td>93</td> <td>93</td> <td>STAN. HEN. AND</td> <td>71</td> <td>93</td> <td>93</td> </tr> <tr> <td>WORD MEAN.</td> <td>70</td> <td>90</td> <td>90</td> <td>WORD MEAN.</td> <td>75</td> <td>103</td> <td>103</td> <td>WORD MEAN.</td> <td>75</td> <td>103</td> <td>103</td> <td>WORD MEAN.</td> <td>75</td> <td>103</td> <td>103</td> </tr> <tr> <td>LONG USAGE</td> <td>66</td> <td>81</td> <td>81</td> <td>LONG USAGE</td> <td>71</td> <td>93</td> <td>93</td> <td>LONG USAGE</td> <td>71</td> <td>93</td> <td>93</td> <td>LONG USAGE</td> <td>71</td> <td>93</td> <td>93</td> </tr> <tr> <td>ARITH. REAR</td> <td>65</td> <td>78</td> <td>78</td> <td>ARITH. REAR</td> <td>68</td> <td>85</td> <td>85</td> <td>ARITH. REAR</td> <td>68</td> <td>85</td> <td>85</td> <td>ARITH. REAR</td> <td>68</td> <td>85</td> <td>85</td> </tr> <tr> <td>LITERATURE</td> <td>62</td> <td>72</td> <td>72</td> <td>LITERATURE</td> <td>69</td> <td>88</td> <td>88</td> <td>LITERATURE</td> <td>69</td> <td>88</td> <td>88</td> <td>LITERATURE</td> <td>69</td> <td>88</td> <td>88</td> </tr> <tr> <td>SOC. ST. I.</td> <td>60</td> <td>68</td> <td>68</td> <td>SOC. ST. I.</td> <td>68</td> <td>85</td> <td>85</td> <td>SOC. ST. I.</td> <td>68</td> <td>85</td> <td>85</td> <td>SOC. ST. I.</td> <td>68</td> <td>85</td> <td>85</td> </tr> <tr> <td>SOC. ST. II</td> <td>62</td> <td>72</td> <td>72</td> <td>SOC. ST. II</td> <td>70</td> <td>90</td> <td>90</td> <td>SOC. ST. II</td> <td>70</td> <td>90</td> <td>90</td> <td>SOC. ST. II</td> <td>70</td> <td>90</td> <td>90</td> </tr> <tr> <td>ELEM. SCI.</td> <td>63</td> <td>74</td> <td>74</td> <td>ELEM. SCI.</td> <td>65</td> <td>78</td> <td>78</td> <td>ELEM. SCI.</td> <td>65</td> <td>78</td> <td>78</td> <td>ELEM. SCI.</td> <td>65</td> <td>78</td> <td>78</td> </tr> <tr> <td>SPELLING</td> <td>71</td> <td>93</td> <td>93</td> <td>SPELLING</td> <td>76</td> <td>106</td> <td>106</td> <td>SPELLING</td> <td>76</td> <td>106</td> <td>106</td> <td>SPELLING</td> <td>76</td> <td>106</td> <td>106</td> </tr> </tbody> </table>						TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	OTIS S.A. TEST	50	13.0	11-4	OTIS S.A. TEST	36	11-4	15-10	ALC. PAPER	76	70	63	ALC. PAPER	76	70	63	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TRAILER - TOT	76	79	58	TRAILER - TOT	91	96	65	TRAILER - TOT	91	96	65	TRAILER - TOT	91	96	65	FORM 1 - COMP	30	30	57	FORM 1 - COMP	40	40	67	FORM 1 - COMP	40	40	67	FORM 1 - COMP	40	40	67	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	STAN. HEN. AND	66	81	81	STAN. HEN. AND	71	93	93	STAN. HEN. AND	71	93	93	STAN. HEN. AND	71	93	93	WORD MEAN.	70	90	90	WORD MEAN.	75	103	103	WORD MEAN.	75	103	103	WORD MEAN.	75	103	103	LONG USAGE	66	81	81	LONG USAGE	71	93	93	LONG USAGE	71	93	93	LONG USAGE	71	93	93	ARITH. REAR	65	78	78	ARITH. REAR	68	85	85	ARITH. REAR	68	85	85	ARITH. REAR	68	85	85	LITERATURE	62	72	72	LITERATURE	69	88	88	LITERATURE	69	88	88	LITERATURE	69	88	88	SOC. ST. I.	60	68	68	SOC. ST. I.	68	85	85	SOC. ST. I.	68	85	85	SOC. ST. I.	68	85	85	SOC. ST. II	62	72	72	SOC. ST. II	70	90	90	SOC. ST. II	70	90	90	SOC. ST. II	70	90	90	ELEM. SCI.	63	74	74	ELEM. SCI.	65	78	78	ELEM. SCI.	65	78	78	ELEM. SCI.	65	78	78	SPELLING	71	93	93	SPELLING	76	106	106	SPELLING	76	106	106	SPELLING	76	106	106
TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile	TEST	Sc.	M.A.	% ile																																																																																																																																																																																																																																							
OTIS S.A. TEST	50	13.0	11-4	OTIS S.A. TEST	36	11-4	15-10	ALC. PAPER	76	70	63	ALC. PAPER	76	70	63																																																																																																																																																																																																																																							
TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile																																																																																																																																																																																																																																							
TRAILER - TOT	76	79	58	TRAILER - TOT	91	96	65	TRAILER - TOT	91	96	65	TRAILER - TOT	91	96	65																																																																																																																																																																																																																																							
FORM 1 - COMP	30	30	57	FORM 1 - COMP	40	40	67	FORM 1 - COMP	40	40	67	FORM 1 - COMP	40	40	67																																																																																																																																																																																																																																							
TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile	TEST	Sc.	P.S.	% ile																																																																																																																																																																																																																																							
STAN. HEN. AND	66	81	81	STAN. HEN. AND	71	93	93	STAN. HEN. AND	71	93	93	STAN. HEN. AND	71	93	93																																																																																																																																																																																																																																							
WORD MEAN.	70	90	90	WORD MEAN.	75	103	103	WORD MEAN.	75	103	103	WORD MEAN.	75	103	103																																																																																																																																																																																																																																							
LONG USAGE	66	81	81	LONG USAGE	71	93	93	LONG USAGE	71	93	93	LONG USAGE	71	93	93																																																																																																																																																																																																																																							
ARITH. REAR	65	78	78	ARITH. REAR	68	85	85	ARITH. REAR	68	85	85	ARITH. REAR	68	85	85																																																																																																																																																																																																																																							
LITERATURE	62	72	72	LITERATURE	69	88	88	LITERATURE	69	88	88	LITERATURE	69	88	88																																																																																																																																																																																																																																							
SOC. ST. I.	60	68	68	SOC. ST. I.	68	85	85	SOC. ST. I.	68	85	85	SOC. ST. I.	68	85	85																																																																																																																																																																																																																																							
SOC. ST. II	62	72	72	SOC. ST. II	70	90	90	SOC. ST. II	70	90	90	SOC. ST. II	70	90	90																																																																																																																																																																																																																																							
ELEM. SCI.	63	74	74	ELEM. SCI.	65	78	78	ELEM. SCI.	65	78	78	ELEM. SCI.	65	78	78																																																																																																																																																																																																																																							
SPELLING	71	93	93	SPELLING	76	106	106	SPELLING	76	106	106	SPELLING	76	106	106																																																																																																																																																																																																																																							
EXPLANATORY NOTES CONCERNING TEST SCORES AND PERCENTILES	<p>MAJORITY OF ACHIEVEMENT SCORES ARE ABOVE MEDIAN (public school, independent school, local school, etc.)</p> <p>SCHOLASTIC APTITUDE SOMEWHAT ABOVE AVERAGE</p> <p>HIGHEST ACHIEVEMENT IS IN LINGUISTIC SUBJECTS AND IN ECONOMICS, LAW, AND BUSINESS. THERE IS GENERAL AGREEMENT BETWEEN SCHOOL MARKS AND TEST SCORES.</p>																																																																																																																																																																																																																																																					

This space reserved for notes by College Admissions Officer or Employer.

Figure 26. Cumulative Record of Edward S. Mansfield over a Six-Year Period (Front of Card). (This is a simplified SRA Cumulative Record Form prepared by Arthur E. Traxler, published by Science Research Associates, 1947. Used with permission of Science Research Associates.)

NAME			BIRTH DATE		BIRTHPLACE		RELIGION		METHODIST		ADDRESS		TELEPHONE											
Mansfield, Edward Samuel			9-18 34		Jamesville, Florida		Methodist		Methodist		28 Lakeside Ave, Newburgh 14620 W Forest, Newburgh		6 1428 6 1428											
FAMILY			OCCUPATION		EDUCATION		RELIGION		HEALTH		DECEASED DATE													
FATHER HOWARD A.			REAL ESTATE DEALER		2 YEARS COLLEGE BROWN UNIVERSITY		METHODIST		GOOD				Note here anything of significance concerning the family or conditions in the home. In January 1951, the mother was fatally injured in an auto accident. Some two weeks later when the accident occurred, the entire family went during the time. Mother took loss badly with resulting strain on other members of family. Readjustment slow but finally effected.											
MOTHER AUREY S.			HOUSEWIFE				METHODIST		GOOD															
STEP PARENT OR GUARDIAN																								
SIBLINGS							BIRTH DATE																	
WILLIAM JARVIS			ACCOUNTANT		13 A ACCOUNTING UNIV OF MIAMI		4-23-24		EXCELLENT		JANUARY, 1951													
MARY RUTH			STUDENT		HUNTER COLLEGE		11-23-30		GOOD															
YEAR AND AGE			1946-1947		C.A. 12		1947-1948		C.A. 13		1948-1949		C.A. 14		1949-1950		C.A. 15		1950-1951		C.A. 16		1951-1952	
INTELLIGENCE QUOTIENT			BINET III, OTIS III		OTIS 114		OTIS 108		OTIS 113		OTIS 113		OTIS 113		OTIS 113		OTIS 113		OTIS 113		OTIS 113		OTIS 113	
COUNSELOR			MISS WARREN		MR. SNIDER		MR. SNIDER		MR. SNIDER		MR. SNIDER		MR. SNIDER		MR. SNIDER		MR. SNIDER		MR. SNIDER		MR. SNIDER		MR. SNIDER	
ATTENDANCE			ABSENT TWO WEEKS FOR TONSILLECTOMY. A LITTLE SLOW IN ADJUSTING TO JR. H.S. PROGRAM - NEW SCHOOL		REGULAR		REGULAR		REGULAR		REGULAR		REGULAR		REGULAR		REGULAR		REGULAR		REGULAR		REGULAR	
ACADEMIC DISCIPLINE			DRAMATICS CLUB		GOOD		GOOD		GOOD		GOOD		GOOD		GOOD		GOOD		GOOD		GOOD		GOOD	
INTERESTS			BOY SCOUTS		DRAMATICS CLUB		BOY SCOUTS		BOY SCOUTS		BOY SCOUTS		BOY SCOUTS		BOY SCOUTS		BOY SCOUTS		BOY SCOUTS		BOY SCOUTS		BOY SCOUTS	
OUT OF SCHOOL			NONE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE		NEWSPAPER DELIVERY ROUTE	
VOCATIONAL EXPERIENCES			DID WELL IN PARTIALITY GIVEN BY DRAMATICS CLUB IN ASSEMBLY		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB		VERY GOOD WORK IN DRAMATICS CLUB	
NOTABLE ACCOMPLISHMENTS			QUITE SUSCEPTIBLE TO GOOD FOLLOWING TONSILLECTOMY-RECOVERY RATHER SLOW		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE		HADN UNUSUALLY IMPROVED OVER LAST YEAR RATHER ACTIVE	
HEALTH AND PHYSICAL VIGOR			NO PLANS AS YET		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL		NO PLANS BEYOND HIGH SCHOOL	
EDUCATIONAL AND VOCATIONAL PLANS			HAS REPORT PENDING RATHER ARGUMENT		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR		HAS REPORT PENDING FOR WORK THIS DEPENDABLE THEN LAST YEAR	
RESPONSIBILITY DEPENDABILITY			SOME ORIGINALITY BUT TENDS TO NEED DIRECTION		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK		HAS SOME ORIGINALITY AND SOME IMAGINATION IN CLUB WORK	
CREATIVENESS AND IMAGINATION			NEUTRAL		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE		GENUALLY CONSTRUCTIVE	
INFLUENCE			HAS ROUTED TO NEW GROUP SLOWLY		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR		GETS ALONG WITH CLASS - MAKES FEWLY WELL. LESS RETIRING THAN LAST YEAR	
SOCIAL ADJUSTMENT			NEED TO VISIT MOTHER TO SECURE COOPERATION TOWARD BETTER SCHOOL ADJUSTMENT		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL		VISITED HOME PARENTS QUITE COOPERATIVE TOWARD SOCIAL BEHAVIOR IS SHOWING MORE SATISFACTORY ADJUSTMENT TO SCHOOL	
COUNSELOR'S NOTES																								
REMARKS																								

EDWARD'S DECISION TO CONTINUE HIS EDUCATION WITH ASSOCIATION FOR ACHIEVEMENT FOR ACHIEVEMENT SEEMS WISE IN LINE WITH HIS ABILITIES AND ACCOMPLISHMENTS. HE DEVELOPED GREATLY DURING SENIOR HIGH SCHOOL AND IS NOW A RESPONSIBLE, WELL RESPECTED BOY

FIGURE 27. Cumulative Record of Edward S. Mansfield over a Six-Year Period (Back of Card).

INTRODUCTION TO TESTING

been good. The family expects to provide financial support to Edward as long as he is in college.

Personally, Edward is a very likable boy. He is a good conversationalist and is well liked by his classmates. He has considerable interest and skill in athletics. He was somewhat slow to develop initiative but has accepted responsibility well during the last years of school. His work in dramatics and debating has been outstanding.

He has good linguistic aptitude and should be able to do acceptable college work if he applies himself conscientiously. He has qualities which should make for success in his chosen profession of law.

Such is the story of the development of Edward Mansfield during his junior- and senior-high-school years. The record form and other personal data collected at the Newtonville public schools should provide valuable information for any college admission officer, teacher, or adviser who consults it. If such a record were also maintained throughout college, the long-term picture would be of real value to any prospective employer.

Note that test results are an important part of

the record kept for this pupil but they are only a part. They have been considered in relation to the performance of Edward's schoolmates and in relation to the results of larger norm groups of public-school pupils. They have been related to all available information about Edward's personality, likes and dislikes, goals, and background. The test results have supplemented, not replaced, the considered judgments of teachers and counselors based on daily contacts. Edward was not a problem pupil in any sense of the term. The uses of his test scores might have been more dramatic if he had shown some distinct disabilities or personality difficulties. But like any other boy he has had certain minor adjustment problems which are interesting if not exceptional. We hope that this discussion has helped to illustrate and clarify some of the contributions of objective testing, a procedure as valuable for the average as for the unusual pupil. For pupils at all levels, measurement helps to lend definiteness and confidence to teaching and guidance.

SUGGESTIONS FOR FURTHER READING

1. *Guidance in Public Secondary Schools*, A Report of the Public School Demonstration Project in Educational Guidance, Educational Records Bulletin No. 28, New York, Educational Records Bureau, October, 1939, pp. 45-73, 203-295.
2. Learned, William S., and Hawkes, Anna L. Rose, *An Experiment in Responsible Learning*, The Carnegie Foundation for the Advancement of Teaching, Bulletin No. 31, New York, The Carnegie Foundation for the Advancement of Teaching, 1940; pp. 36-61.
3. Super, Donald E., *Appraising Vocational Fitness*, New York, Harper & Brothers, 1945, pp. 628-642.
4. Wood, Ben D., and Haefner, Ralph, *Measuring and Guiding Individual Growth*, New York, Silver Burdett Company, 1948, pp. 65-69, 437-443, 470-478.

Appendix

MAJOR TEST PUBLISHERS

- Bureau of Educational Research and Service, State
University of Iowa
Iowa City, Iowa
- Bureau of Publications, Teachers College, Colum-
bia University
New York 27, New York
- California Test Bureau
5916 Hollywood Boulevard, Los Angeles 28, Cali-
fornia
110 South Dickinson Street, Madison 3, Wisconsin
206 Bridge Street, New Cumberland, Pennsyl-
vania
- Educational Test Bureau
720 Washington Avenue, S.E., Minneapolis 14,
Minnesota
3433 Walnut Street, Philadelphia 4, Pennsylvania
2106 Pierce Avenue, Nashville, Tennessee
- Educational Testing Service
20 Nassau Street, Princeton, New Jersey
4641 Hollywood Boulevard, Los Angeles 28, Cali-
fornia
- Houghton Mifflin Company
2 Park Street, Boston 7, Massachusetts
432 Fourth Avenue, New York 16, New York
2500 Prairie Avenue, Chicago 16, Illinois
500 Howard Street, San Francisco 5, California
- 715 Browder Street, Dallas 1, Texas
39 Harris Street, Atlanta 3, Georgia
- The Psychological Corporation
522 Fifth Avenue, New York 18, New York
- Public School Publishing Company
509-513 North East Street, Bloomington, Illinois
- Science Research Associates
57 West Grand Avenue, Chicago 10, Illinois
- Stanford University Press
Stanford University, California
- Steck Company
Austin, Texas
- C. H. Stoelting Company
424 North Homan Avenue, Chicago 24, Illinois
- University of Minnesota Press, University of Min-
nesota
Minneapolis 14, Minnesota
- World Book Company
Yonkers 5, New York
2126 Prairie Avenue, Chicago 16, Illinois
6 Beacon Street, Boston 8, Massachusetts
441 West Peachtree Street, N.E., Atlanta 3,
Georgia
707 Browder Street, Dallas 1, Texas
121 Second Street, San Francisco 5, California

Index

- Ability, *see* Aptitude
- Achievement, 5
 meaning of, 8
 measurement of, 8-9
 objective analysis of, 8
 subjective analysis of, 8
 test results illustrated, 72, 76-80
- Adjustment, measurement of personal and social, 9-10
- Administration of tests, 30-34
 accuracy of, 32, 34, 91
 criteria for examiner in the, 31-33
 ease of, 25
 general principles of, 30-31
 information for students about, 33
 preparing students for, 33, 91
- Administrative uses of tests, 55, 93
- American Council on Education
Psychological Examination, 64-66, 81-83, 103
- American Institute of Accountants, Committee on Selection of Personnel, 95
- Analysis of test results, *see* Interpretation and analysis of test records
- Application of test results, 96-108
 in action research, 93
 in aiding individual pupils, 93
 in appraisal of effectiveness of instruction, 93
 in diagnostic measurement, 93
 in guidance, 92
 in parent counseling, 94
 specific illustrations of, 96-108
 See also Guidance
- Aptitude, meaning of, 5
 measurement of general, 6-7, 18
 special, 7, 18
 See also Testing program, minimum and "maximum"
- Bell, Bernard I., *quoted*, 2
- Central tendency, measures of, 52
- Class lists, 42, 62, 64-69
 illustrated, 65
 inefficiency of, as permanent records, 68
- Commission on Youth Problems, report of, 2
- Cooperative Achievement Tests*, distributions based on, 40, 48-49, 57, 58
 profile form for, 69, 74
 record form illustrated, 72
- Cooperative English Test*, timing of, 32
- Correlation, coefficient of, 20-24, 54-56
 interpretation of, 55
 meaning of, 54
 negative, 55
 probable error of, 56
 value of, 55
 See also Reliability; Validity
- Cost of tests, 25
- Diagnostic Reading Tests, Survey Section*, 65, 93
- Diagnostic tests, 18, 64, 93
- Differential Aptitude Tests*, profile for, 69
 profile illustrated, 70
- Distribution of scores, 40, 48-49, 57, 58
 illustrated, 41, 48, 49, 57, 58, 59
 uses of, 58, 59
 See also Normal curve
- Educational quotient, 46
- Educational Records Bureau, cumulative record card, illustrated, 82-86
 cumulative record card prepared jointly with Secondary Education Board, illustrated, 78-80
 report from files of, 64-65
 test selection procedure, 13-14
 testing program of, 17-19
- Essay-type examinations, 8
- Guidance, 6, 15, 32, 71, 93, 94
 occupational and vocational, 9, 18, 71, 103, 108
 See also Application of test results
- Guiterman, Arthur, *quoted*, 1
- Intelligence quotient, 46, 53, 92, 96-97, 99
- Intelligence testing, 17
 See also Testing program, minimum and "maximum"
- Interests, 5
 in minimum testing program, 16
 in ultimate or "maximum" testing program, 18
 measurement of, 9
 results of tests of, illustrated, 73
 vocational, 9, 18, 71, 103, 108

INDEX

- Interpretation and analysis of
test records, 56-66
for group, 57-66
for individual, 59-62
use of statistics in, 47-54, 56, 91
See also Records
- Interpretation and analysis of
test results, 45-67, 96-108
example of in specific case,
96-108
norms, *see* Norms
purpose of statistics in, 40, 47
study of answers in test book-
lets for, 56
subjectivity involved in, 11
teacher training for, 91
- Interquartile range, 51-52
- Iowa Silent Reading Test*, 59, 63,
64
class record for, 63
publisher's distribution sheet
for, 59
- Kuder Preference Record*, 71
profile for, 73
- Kuder-Richardson method of de-
riving reliability, 23
- Locke, John, *quoted*, 1-2
- Mean, 48, 50
- Median, 40, 48, 49, 50
- Mental ability, *see* Aptitude
- Mental Measurements Year-
books*, 20, 27
- Mental testing, *see* Aptitude; In-
telligence testing
- Metropolitan Achievement Test*,
77
- Normal curve, defined, 49
illustrated, 50, 51, 54
- Norms, adequacy of, 26, 42
age, 46
caution in using, 47
educational quotients as, 46
grade, 46
independent-school, 42
meaningfulness of, 38
modal-age, 46
- Norms—(*Continued*)
national, 58
occupational group, 26
percentile, 45-46, 53-54, 66, 69
regional, 26
selection of, 26
sex group, 26
- Objective tests, 5-12
degree of objectivity in, 24
item analysis to obtain greater
objectivity in, 24
limitations of, 11
- Objectives of testing, 5-6, 14
- Otis Self-Administering Test of
Mental Ability*, 96-97
equivalent I.Q.'s derived from,
for the American Council
on Education Psychological
Examination, 64
- Percentiles, 45-46, 53-54, 66, 69
- Personality, limitations of testing
in, 10
tests of, 9-10
See also Testing program,
minimum and "maxi-
mum"
- Printing and format of tests, im-
portance of, 26
- Probable error, 56
functions of, 56
- Projective tests, 10
- Psychological Abstracts*, test re-
search summarized in, 28
- Quartiles, 40, 51-52, 57
- Reading tests, 16, 17, 18
See also Testing program,
minimum and "maxi-
mum"
- Recording of test results, 69-88,
91
illustrated, 63, 65, 72, 73, 78-
80, 82-86, 98, 100-102, 104-
107
See also Records
- Records, anecdotal, 19
cumulative, 66, 69, 74-87; il-
lustrated, 78-80, 82-86, 97-
108; principles for selecting,
74-75
individual, 62; folders for, 69;
limitations of, 71, 74
profiles of test, 69; illustrated,
70, 72, 73
- Regression effect, 66
- Reliability, coefficient of, 22-23
Kuder-Richardson, 23
meaning of, 21
method of deriving, by com-
parable forms, 22
minimum required for group
prediction, 23
minimum required for individ-
ual diagnosis, 23
Spearman-Brown, 23
test-retest, 22
See also Correlation
- Reliability of prediction, 55
standard error of measurement
as indication of, 24
- Reports, use of test results in, to
colleges, 94; to employers,
94
- Scholastic Aptitude Test*, Col-
lege Entrance Examination
Board, 92
- Scores, range of, 48
raw, 44
Scaled, 24, 53
- Scoring, accuracy of, 91
advantages of, by an agency,
36
advantages of, by teachers, 36-
37
by an agency, 36, 37-42
by hand, 38
ease of, 25
machine, 33-34, 40, 91
pupil self-, 26
teacher, 36-37
- Secondary Education Board, cu-
mulative record card pre-
pared jointly with Educa-

INDEX

- Sec. Educ. Board—(*Continued*)
 tional Records Bureau, illustrated, 78-80
Junior Scholastic Aptitude Test, 92
 Spearman-Brown method of deriving reliability, 23
 SRA cumulative record form, illustrated, 100-101, 104-107
 Standard deviation, 24, 52, 54
 Standard error of measurement, 24
Stanford Achievement Test, 97, 99
 profile for, illustrated, 98, 102
Stanford-Binet Scale, 17, 96, 97
 Statistics, minimum essentials needed in test interpretation, 47-55
- Test construction, teacher cooperation in, 89
 Test publishers, list of, 109
 Test selection, Educational Records Bureau procedure in, 13-14
 Testing program, 13-19
 cost of, 25
 experimental part of, 18
 minimum, 16-17
 planning of, 13-15, 90
 schedule for, 30-31
 selection of tests for, 26-27, 90
 sources of information regarding, 27-28
 ultimate or "maximum," 17-18
 Tests, desirable characteristics of, 20-26
- Traxler Silent Reading Test*, 76, 96
 Use of test results, 89-95
 prerequisites to, 89-92
 See also Interpretation and analysis of test records
 Validity, correlations to determine, 21
 curricular, 21
 derivation of, 20-21
 meaning of, 20
 Variability, 52
 Vocational guidance, *see* Guidance
Yale Educational Aptitude Tests, 93

